



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

INSTITUTO DE BIOTECNOLOGÍA

Departamento de Ingeniería Celular y Biocatálisis

Análisis del Origen y Evolución del Metabolismo de Nucleótidos Desde una Perspectiva Genómica

T E S I S
QUE PARA OPTAR POR EL GRADO DE:
Doctor en Ciencias

P R E S E N T A:

M.C. Dagoberto Armenta Medina

TUTOR PRINCIPAL
Dr. Ernesto Pérez Rueda
Instituto De Biotecnología

MIEMBROS DEL COMITÉ TUTOR
Dr. Arturo Becerra Bracho, Facultad de Ciencias
Dr. Maximino Aldana Gonzales, Instituto de Ciencias Físicas

Cuernavaca, Morelos.

Marzo, 2015.



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.

AGRADECIMIENTOS:

A CONACYT por la beca que me fue otorgada durante la realización del posgrado de doctorado en ciencias bioquímicas.

A la unidad de docencia por su disponibilidad y eficiencia en los trámites realizados durante mi estancia en el instituto de biotecnología.

A mi comité tutorial por contribuir en el desarrollo crítico de mi trabajo de tesis.

A los miembros de mi jurado de examen por las observaciones constructivas realizadas al presente trabajo.

A todos los miembros y ex miembros del grupo de Lorenzo y Ernesto que he tenido suerte de conocer. Por su contribución académica y su amistad.

A todos mis amigos, con fortuna son bastantes y no me gustaría dejar alguno fuera. Todos en conjunto han aportado elementos muy valiosos dentro de mi desarrollo humano. Gracias por su invaluable apoyo. A Paulina por mi aprendizaje y desarrollo personal.

Al Dr. Lorenzo Segovia por el espacio y el buen ambiente generado en su grupo. Al igual que otros como el Dr. Lazcano, Becerra, han promovido el enfoque evolutivo en México del cual ha sido de gran influencia en mi estudio.

Al Dr. Ernesto Pérez Rueda el cual ha sido un gran facilitador en mi desarrollo científico. Con su entusiasmo y atención siempre pendiente a mis necesidades.

A toda mi familia especialmente a mis padres y hermanos, cualquier descripción sobre su amor, afecto y apoyo quedaría corto.

El presente trabajo se realizó bajo la dirección del Dr. Ernesto Pérez Rueda, en el instituto de Biotecnología de la Universidad Nacional Autónoma de México. Durante la realización de mismo se recibió una beca del Consejo Nacional de Ciencia y Tecnología (CONACYT).

COMITÉ TUTORAL:

Dr. Ernesto Pérez Rueda (IBT-UNAM)

Dr. Arturo Becerra Bracho (FC-UNAM)

Dr. Maximino Aldana Gonzales (ICF-UNAM)

JURADO DE EXAMEN:

Dr. León Martínez Castilla (FQ-UNAM)

Dr. Guillermo Gosset Lagarda (IBT-UNAM)

Dr. Rosa María Gutiérrez Ríos (IBT-UNAM)

Dr. Daniel Segura Gonzales (IBT-UNAM)

Dr. Armando Hernández Mendoza (FC-UAEM)

Seek to understand the root.

It is futile to argue as to which single leaf, which design of branch, or which attractive flower you like; when you understand the root, you understand all its blossoming.

B. Lee

INDICE

ABSTRACT.....	1
RESUMEN.....	2
1- INTRODUCCIÓN	3
2- ANTECEDENTES	6
3- PLANTEAMIENTO DEL PROBLEMA	8
4- JUSTIFICACIÓN	9
5- HIPÓTESIS	10
6- OBJETIVO GENERAL.....	11
7- OBJETIVOS PARTICULARES	11
8- MATERIAL Y METODOS.....	12
9-RESULTADOS Y DISCUSIÓN	14
9.1- Distribución taxonómica de las enzimas del metabolismo de Nucleótidos	14
9.2- Evolución del metabolismo de Purinas	16
9.2.1- Biosíntesis <i>de novo</i> de purinas	16
9.2.2- Vía de salvamento de Purinas	25
9.2.3- Integrando el análisis de las rutas de salvamento y <i>de novo</i> de purinas.	30
9.3- Evolución del Metabolismo de Pirimidinas	34
9.3.1- Biosíntesis <i>de novo</i> de pirimidinas.....	34
9.3.2- Rutas de salvamento Pirimidinas	39
9.3.3- Integrando el análisis evolutivo de las vías <i>de novo</i> y salvamento del metabolismo de pirimidinas.....	42
10- CONCLUSIONES	44
11- PERSPECTIVAS	47
12- REFERENCIAS.....	48
13-APÉNDICE.....	53
13.1- Distribución taxonómica en la biosíntesis de purinas (S1)	53
13.2- Distribución taxonómica en la biosíntesis de pirimidinas (S2).....	54
13.3- Similitud estructural entre la Inosina y Guanina (S3).....	55
13.4- Glosario.....	55
14-ANEXO	59
14.1 Artículo publicado en BMC Genomics	59

ABSTRACT

Nucleotide metabolism is central to all biological systems, due to their essential role in genetic information and energy transfer, which in turn suggests its possible presence in the last common ancestor (LCA) of *Bacteria*, *Archaea* and *Eukarya*. In this context, elucidation of the contribution of the origin and diversification of *de novo* and salvage pathways of nucleotide metabolism in the evolution of organisms will allow us to understand the links between the enzymatic steps associated with the LCA and the emergence of the first metabolic pathways. In this work, the taxonomical distribution of the enzymes associated with nucleotide metabolism was evaluated in 1,606 complete genomes. From these analyses, it was suggested that divergence at the enzymatic level correlates with environmental changes. In addition, segments of salvage and *de novo* pathways were probably complementary in the LCA to the synthesis of purines and pyrimidines. We also suggest that a large portion of the pathway to inosine 5'-monophosphate (IMP) in purines could have been involved in thiamine synthesis or its derivatives in early stages of cellular evolution, correlating with the fact that these molecules may have played an active role in the protein-RNA world. The analysis presented here provides general observations concerning the adaptation of the enzymatic steps in the early stages of the emergence of life and the LCA.

RESUMEN

El metabolismo de nucleótidos es fundamental para todos los sistemas biológicos, debido a su papel esencial en la transferencia de información genética y de energía, los cuales a su vez sugieren su posible presencia en el último ancestro común (LCA) de bacterias, arqueas y eucariotes. En este contexto, la elucidación de la contribución del origen y diversificación de las rutas de novo y salvamento del metabolismo de nucleótidos en la evolución de los organismos nos permitirá comprender los vínculos entre los pasos enzimáticos asociados al LCA y el surgimiento de las primeras vías metabólicas. En este trabajo, se evaluó la distribución taxonómica de las enzimas asociadas al metabolismo de nucleótidos en 1606 genomas completos. A partir de este análisis, se encontró que la divergencia a nivel enzimático correlaciona con los cambios ambientales. Además, los segmentos de las rutas de salvamento y de novo fueron probablemente complementarias en la LCA para la síntesis de purinas y pirimidinas. También se sugiere que una gran parte de la vía a inosina 5'-monofosfato (IMP) en purinas podría haber estado involucrada en la síntesis de tiamina o sus derivados en los primeros estadios de la evolución celular, correlacionando con el hecho de que estas moléculas pudieron haber desempeñado un papel activo en el mundo proteína - ARN. El análisis presentado en este trabajo proporciona observaciones generales relativas a la adaptación de primeros pasos enzimáticos en el LCA.

1- INTRODUCCIÓN

El metabolismo representa un proceso global a través del cual los sistemas vivos adquieren y utilizan energía para realizar sus diferentes funciones. Este proceso se realiza acoplando las reacciones exergónicas de oxidación de nutrientes a los procesos endergónicos necesarios para mantener el estado vital de la células, tales como el trabajo mecánico, el transporte activo y la biosíntesis de moléculas¹. En este sentido, el metabolismo es considerado uno de los procesos biológicos más antiguos, el cual pudo haber evolucionado como consecuencia de las presiones ocasionadas por el agotamiento de factores vitales en la sopa primitiva, referida en la mayormente aceptada teoría heterotrófica del origen de la vida²⁻⁴.

De acuerdo con lo anterior, el estudio del origen y evolución del metabolismo representa uno de los principales retos para el entendimiento de las etapas tempranas de la emergencia de la vida y contribuye en la comprensión de los procesos de adaptación de la vida celular, la emergencia de altos niveles de organización (multicelularidad) y del desarrollo de la complejidad de los seres vivos, entre otros⁵. En este sentido una pregunta que se ha planteado desde hace más de seis décadas es ¿cómo ha evolucionado el metabolismo hasta llegar a su estado actual?. Al respecto se han planteado dos modelos principales de evolución, el denominado paso a paso o *stepwise*⁶ y el de mosaico o *patchwork*⁷. El modelo de paso a paso sugiere que en respuesta al agotamiento de sustratos en el medio, un gen X existente se duplicó y posteriormente se especializó en una nueva reacción, conservando la especificidad por sustratos similares en pasos enzimáticos consecutivos. Por contraparte, el modelo de mosaico propone que el metabolismo es el resultado del reclutamiento de enzimas con baja especificidad, que pueden reaccionar con una gran cantidad de sustratos

químicamente relacionados, pudiendo mantener en los eventos de duplicación y divergencia tipos de reacciones similares sin necesidad de estar a pasos enzimáticos consecutivos especializándose en alguno de sus múltiples sustratos originales, Figura 1. De manera adicional, distintos grupos han propuesto otros modelos, dentro de los que sobresalen la propuesta por Lazcano-Miller² sobre el origen semi-enzimático de las rutas metabólicas. Este modelo introduce el concepto de enzimas iniciadoras o “*starter types enzymes*” (Figura 1), donde se sugiere que estas enzimas surgieron a partir de reacciones lentas, no enzimáticas, las cuales se fueron perfeccionando a través de su interacción con proteínas a lo largo del tiempo. Uno de los ejemplos más representativos es el de las transaminaciones catalizadas por piridoxal, que se sugiere dieron lugar a enzimas semillas a partir de las cuales se generaran variantes por eventos de duplicación y divergencia, permitiendo la evolución del metabolismo⁸.

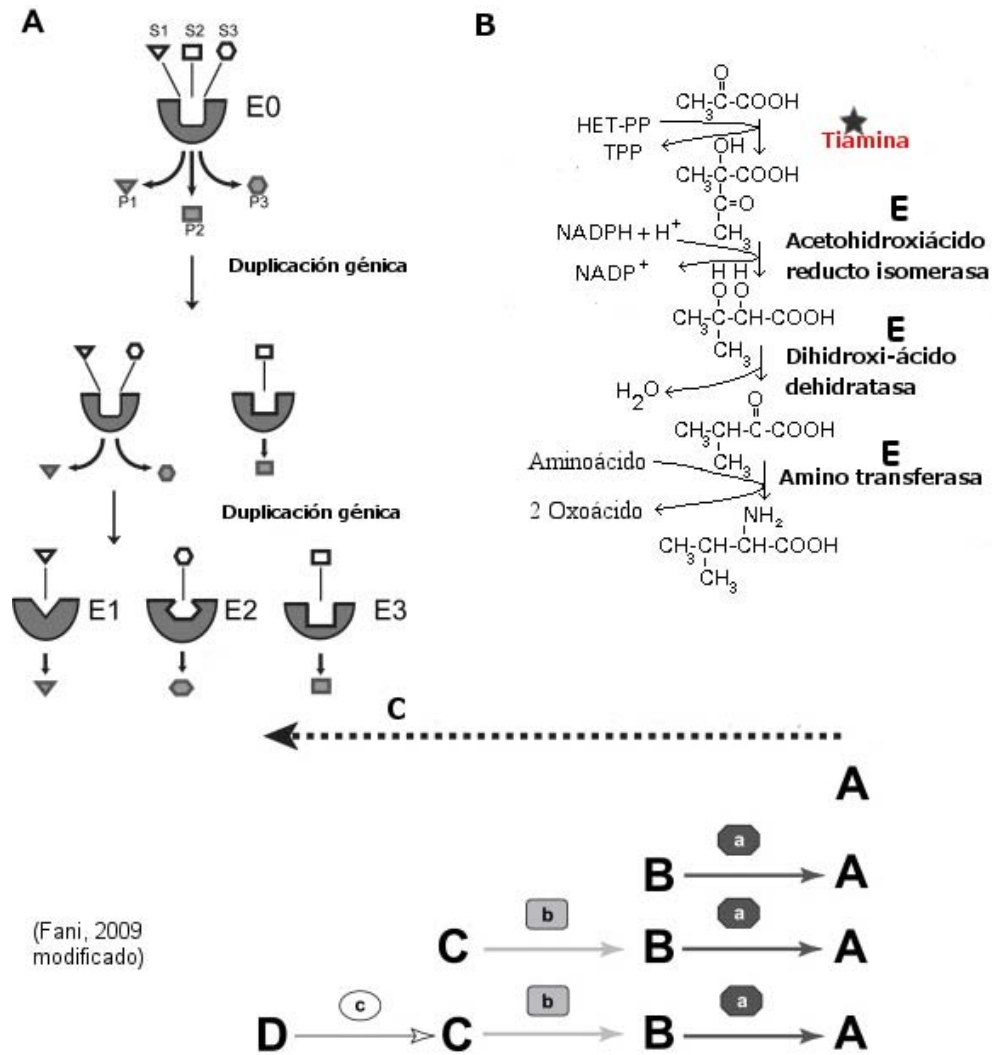


Figura.1- Modelos de evolución del metabolismo. a) Mosaico, las enzimas son señaladas con la letra E; b) semi-enzimático, donde las enzimas son señaladas con la letra E y la parte no enzimática con una estrella. c) paso a paso, las enzimas se describen en letras mayúsculas y los sustratos en letras minúsculas.

La mayoría de los análisis de la evolución del metabolismo apoyan, en mayor medida, la hipótesis de mosaico. En este modelo de evolución se sugiere que un pequeño número de enzimas con amplia especificidad existió en las primeras etapas de la evolución metabólica y que los genes que codifican estas enzimas se fueron especializando a través de procesos

de duplicación y divergencia⁹. Aunque en la mayoría de los estudios sobre la evolución del metabolismo predomine la hipótesis de mosaico, también se han evidenciado ejemplos que apoyan los otros modelos, sugiriendo que estos pueden ser complementarios^{10,11}.

2- ANTECEDENTES

Debido a la gran cantidad de estudios bioquímicos generados en las últimas décadas, ha sido posible el estudio integrativo del metabolismo desde una perspectiva de redes, donde los nodos representan sustratos o enzimas y las conexiones representan las relaciones entre ellos¹¹⁻¹³. Desde esta perspectiva, el estudio de las redes metabólicas se centra en la descripción de las propiedades topológicas, tales como la existencia de módulos funcionales, dando especial relevancia a la agrupación y la formación de motivos que muestra la existencia de atributos similares a las redes de mundo pequeño y de libre escala^{12,14}. De igual forma, el desarrollo de tecnologías como la secuenciación masiva de genomas completos ha hecho posible el estudio comparativo del metabolismo a escalas genómicas, permitiendo evaluar su origen y evolución en los seres vivos^{5,10,15-18}.

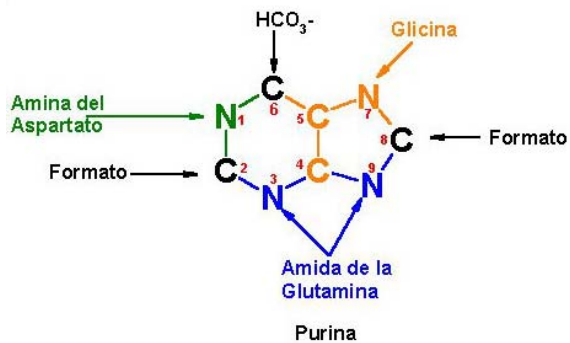
En trabajos recientes, la presencia en los tres dominios celulares de algunas vías y ramas, tales como las asociadas a la biosíntesis de aminoácidos, sugiere su presencia en el último ancestro común (LCA) de Bacterias, Arqueas y Eucariotes^{18,19}. Lo anterior se fundamenta en la descripción monofilética del LCA, la cual es apoyada por árboles tanto de RNA ribosomal, como de contenido génico^{20,21} y los rasgos fundamentales comunes entre los seres vivos de los tres dominios celulares, tales como el código genético y las características básicas de la maquinaria de expresión génica¹⁹.

En este sentido, a pesar de que existen contribuciones acerca de la evolución del metabolismo, pocos trabajos han abordado el estudio del metabolismo de nucleótidos mediante enfoques de redes y genómica comparativa a gran escala, a pesar de su gran relevancia en todos los organismos vivos^{22,23}. El metabolismo de nucleótidos es central en los sistemas biológicos debido a su papel en la transferencia de la información genética y energía. De hecho, se ha descrito como uno de los metabolismos más antiguos. En particular, la aparición del plegamiento del tipo “P-loop hidrolasa” asociado a este metabolismo refuerza la noción de su antigüedad⁵. El plegamiento P-loop hidrolasa se encontró ser de los más abundantes en los seres vivos estando especialmente ampliamente distribuido en el metabolismo de nucleótidos. En este sentido las filogenias de plegamientos que sugieren la antigüedad del plegamiento p-loop hidrolasa podrían estar reflejando de igual forma la antigüedad del metabolismo de nucleótidos. De manera adicional, muchos de los intermediarios asociados a este módulo metabólico han sido íntimamente asociados con la química prebiótica y cercanos al origen de la vida^{22,24,25}. Lo anterior denota la importancia de contribuir en el entendimiento del origen y evolución del metabolismo de nucleótidos, desde una perspectiva de genómica comparativa.

3- PLANTEAMIENTO DEL PROBLEMA

Debido a que el metabolismo de nucleótidos es fundamental para los seres vivos y podría estar asociado a etapas tempranas de la evolución celular, consideramos indispensable su estudio de forma exhaustiva. Desde esta forma, para el estudio del metabolismo de nucleótidos se utilizó un enfoque de redes y de genómica comparativa que permitió su reconstrucción y análisis. Este análisis involucró la evaluación de la distribución taxonómica de sus familias enzimáticas en los tres dominios celulares, en conjunto con evidencias teóricas y experimentales. Con base en este análisis, fue posible determinar la contribución de las vías de biosíntesis *de novo* y de salvamento de ribonucleótidos y desoxirribonucleótidos en el LCA, Figura 2. Adicionalmente, debido a la antigüedad de este metabolismo, su análisis nos permitirá hacer una conexión entre el LCA y los primeros pasos en el diseño de las redes metabólicas²⁶⁻²⁸. Por último, el estudio del metabolismo de nucleótidos nos permitirá correlacionar sus transformaciones derivadas de procesos de adaptación celular.

A) Rutas de Novo



B) Rutas de Salvamento

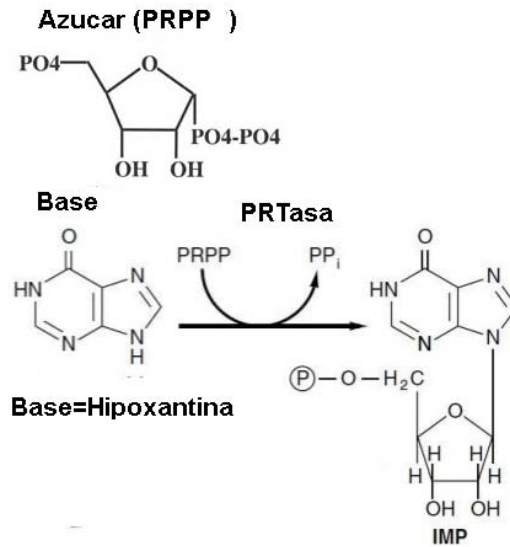


Figura.2- Rutas *de novo* y de salvamento de nucleótidos. A) En las rutas *de novo* se van adhiriendo moléculas precursoras mediante un serie de pasos enzimáticos al azúcar (PRPP). B) Rutas de salvamento, las bases preformadas se adhieren al PRPP mediante la enzima PRTasa. Esta reacción se puede dar para las diferentes bases nitrogenadas.

4- JUSTIFICACIÓN

El estudio del origen y evolución del metabolismo de nucleótidos es de gran importancia, ya que su análisis contribuye a entender diversos aspectos, tales como los siguientes:

a) Debido a que las vías de biosíntesis *de novo* y de salvamento de ribonucleótidos y desoxirribonucleótidos son descritas como de las más antiguas, su estudio contribuirá en el entendimiento de las etapas tempranas de la evolución celular.

b) Cual es la contribución de las vías de biosíntesis y de salvamento en el LCA para la síntesis de nucleótidos.

5- HIPÓTESIS

Por medio de genómica comparativa y el uso de perfiles filogenéticos será posible evaluar la antigüedad del metabolismo de nucleótidos y sus transformaciones derivadas de procesos de adaptación.

6- OBJETIVO GENERAL

Analizar el origen y evolución del metabolismo de nucleótidos utilizando genómica comparativa y perfiles filogenéticos en los 3 dominios celulares de la vida.

7- OBJETIVOS PARTICULARES

- a) Analizar la distribución taxonómica de las enzimas involucradas en la biosíntesis de nucleótidos.
- b) Proponer la posible capacidad de biosíntesis de los ribonucleótidos y desoxiribonucleótidos en el LCA de acuerdo a la distribución taxonómica de sus enzimas en los 3 dominios celulares.
- c) Correlacionar las transformaciones del metabolismo de nucleótidos con cambios ambientales y procesos de adaptación celular.

8- MATERIAL Y MÉTODOS

Perfiles (RPS-Blast)

Con el fin de seleccionar los números enzimáticos y sus correspondientes enzimas pertenecientes al metabolismo de nucleótidos, las bases de datos Enciclopedia de Genes y Genomas de Kyoto (KEGG)²⁹ y la base de datos universal de vías metabólica (MetaCyc)³⁰ fueron exhaustivamente exploradas. En total, se colectaron 120 números enzimáticos según KEGG y sus correspondientes enzimas asociadas de acuerdo a la base de datos Enzyme. En un segundo paso, se utilizaron perfiles de matrices de puntuación posición específica (pssm) para buscar la presencia de miembros de familias de enzimas en genomas completos mediante reversed position specific BLAST (RPS- BLAST). Estos perfiles fueron extraídos de PRIAM, una base de datos especializada para la detección de secuencias enzimáticas³¹. La base de datos PRIAM integra características representativas de alineamientos de familias enzimáticas asociadas a diferentes números enzimáticos, incluyendo en cada alineamiento a los miembros de una familia enzimática en particular, de acuerdo a la base de datos Enzyme³². En este sentido, un paso enzimático puede estar asociado a más de un perfil, tal como los complejos de proteínas, o familias de proteínas no homólogas (enzimas análogas) o bien, los pasos enzimáticos llevados a cabo por diferentes familias enzimáticas, pero que aún mantienen relaciones de homología (por ejemplo, parálogos). Cabe mencionar que los perfiles fueron manualmente curados basándose en los dominios funcionales anotados según Enzyme DB³², Figura 3.

Función Enzimática

Para el análisis de las funciones enzimáticas, el mejor hit para cada secuencia, con un valor E de $\leq 10^{-10}$ y una cobertura de $\geq 55\%$ con relación al perfil, fue considerado. Criterios

similares se han descrito anteriormente para la anotación enzimática de genomas completos³³, Figura 3.

Distribución Taxonómica

En total, 151 perfiles de secuencia asociados a 120 reacciones enzimáticas relacionadas con el metabolismo de nucleótidos se evaluaron en 2044 genomas completos, basado en la comparación de perfiles, utilizando RPS- BLAST³⁴. Los organismos clasificados como parásitos obligados o aquellos organismos con un genoma reducido, es decir, menos de 1000 genes, no se consideraron en este estudio, con el objetivo de excluir a un posible sesgo asociado con la pérdida masiva de genes, como se ha descrito anteriormente^{17,19}, lo cual nos dejó con un total de 1606 de genomas completos. Con el fin de excluir la redundancia en los genomas analizados, los organismos fueron agrupados con base en sus clasificaciones taxonómicas, con el fin de obtener una medida normalizada de la distribución taxonómica de las enzimas de acuerdo con los siguientes pasos. En el primer paso, se obtuvo el promedio de presencia/ausencia de enzimas por género, que a su vez se utilizó en una segunda etapa para calcular la media de presencia/ausencia de enzimas por clado. Los clados corresponden a las categorías taxonómicas de acuerdo al Joint Genome Institute's Integrated Microbial Genomics. Finalmente, consideramos como enzimas ampliamente distribuidas aquellas enzimas presentes en más del 50 por ciento de los clados del mismo dominio celular, Figura 3, S1 y S2 (Figuras suplementarias ver índice).

Identificación de dominios estructurales y familias

Los dominios estructurales de proteínas y de sus familias se identificaron utilizando los modelos depositados en la base de datos de Superfamily³⁵. Esta base de datos contiene los modelos ocultos de Markov (HMM) para cada superfamilia, que se clasifican en función de

metabólicas^{9,16,18}. En este sentido, el origen y evolución del metabolismo de nucleótidos fue exhaustivamente explorado en organismos pertenecientes a los tres dominios celulares, bacterias, arqueas y eucariotas, considerando la distribución taxonómica de sus repertorios enzimáticos. De esta forma, cada actividad enzimática codificada por un número EC puede presentar uno o más perfiles asociados, que son vectores individuales de la presencia / ausencia de enzimas en todos los genomas.

En este trabajo, se utilizó un total de 151 perfiles asociados a 120 reacciones enzimáticas relacionadas con el metabolismo de nucleótidos para evaluar su presencia en 1606 genomas mediante el uso de RPS-BLAST (Figuras suplementarias S1 y S2 ver índice). Sobre la base de estas comparaciones, el origen evolutivo del metabolismo de nucleótidos puede ser rastreado, incluso cerca del LCA. En este sentido, se propone que las enzimas ampliamente distribuidas en los tres dominios celulares pudieron haber estado presentes en el LCA^{10,17-20,37}. Alternativamente, la presencia de enzimas restringidas a clados o dominios celulares específicos sugiere ser el resultado de adaptaciones de los organismos o dominios celulares a ambientes y estilos de vidas particulares o novedades metabólicas que favorecen la selección de estos organismos. Con base en estas consideraciones, en las siguientes secciones, se discuten los resultados más significativos.

9.2- Evolución del metabolismo de Purinas

9.2.1- Biosíntesis *de novo* de purinas

La biosíntesis *de novo* de purinas, iniciando de la D-ribosa-1- fosfato hacia la producción de inosina-5'-monofosfato (IMP), el principal intermediario en la biosíntesis de ribonucleótidos y desoxirribonucleótidos, guanidina y adenina, sigue una rama lineal. El primer paso enzimático está asociado a la fosfoglucomutasa, EC 5.4.2.2 o fosfopentomutasa, 5.4.2.7; mientras que el segundo paso se asocia con la familia enzimática ribosa-fosfato difosfocinasa (2.7.6.1). Ambos pasos son necesarios para la síntesis de 5-fosfo-alfa-D-ribosil-1-pirofosfato (PRPP). Con base en su distribución taxonómica, las enzimas asociadas a los números EC 5.4.2.2 y 2.7.6.1 fueron identificadas como ampliamente distribuidas en los tres dominios celulares (Bacteria, Arquea y Eucaria), lo que sugiere la biosíntesis de PRPP en el LCA. De hecho, PRPP es un precursor clave para la biosíntesis *de novo* y salvamento de purinas y pirimidinas, sin embargo, este intermediario es inestable y susceptible a la hidrólisis. Nuestros datos de distribución taxonómica sugieren la existencia de una vía de biosíntesis de PRPP en el LCA, donde es probable que su síntesis abiótica, si es que ocurrió, no lograra su acumulación en suficientes cantidades debido a su inestabilidad química²².

El primer paso para la biosíntesis de purinas; es decir, la catálisis a ribosa-5-fosfato a partir de ribosa 1-fosfato como mencionamos previamente, se logra por cualquiera de las dos familias enzimáticas relacionadas con los números EC 5.4.2.2 y 5.4.2.7. Estas dos familias enzimáticas sugieren ser análogas, ya que no se detectó homología evidente a nivel de familia y superfamilia estructural. La familia enzimática asociada al EC 5.4.2.7 se

distribuye parcialmente en bacterias, principalmente en organismos de vida libre asociados a un hospedero, tales como *Streptococcus pneumoniae* y *Lactobacillus rhamnosus*, sin embargo, está ausente en las arqueas y eucariotes, lo que sugiere su aparición posterior a la divergencia LCA, probablemente como una adaptación secundaria asociada al huésped.

A partir del intermediario PRPP, hacia la biosíntesis de IMP, se identificaron las familias enzimáticas pertenecientes a cinco pasos catalíticos (amido-fosforibosiltransferasa, 2.4.2.14; fosforibosilamin-glicin-ligasa, 6.3.4.13; fosforibosilglicinamide formiltransferasa, 2.1.2.2; fosforibosilformilglicinamidina sintasa, 6.3.5.3; fosforibosilformilglicinamidina ciclo-ligasa, 6.3.3.1) para la transformación de PRPP en AIR. La mayoría de estas enzimas fueron identificadas como ampliamente distribuidas en los tres dominios celulares, lo que sugiere la presencia de esta rama en el LCA (Figura 4). El paso enzimático asociado al EC 2.1.2.2 es responsable de la transformación del ribotido de glicinamida (GAR) a ribotido de formilglicinamida (FGAR) y puede llevarse a cabo por dos familias enzimáticas con diferente origen evolutivo (análogas), PurN (Figura 4 rectángulo color oro) o formiltransferasa fosforibosilglicinamida y PurT o fosforibosilglicinamida formiltransferasa 2. La familia de proteínas PurN utiliza derivados de la biosíntesis de folatos como sustratos. Esta familia fue identificada como ampliamente distribuida en bacterias y eucariotes y parcialmente en arqueas. Alternativamente, la familia de proteínas PurT se encuentra parcialmente distribuida en arqueas y bacterias y escasamente en eucariotes.

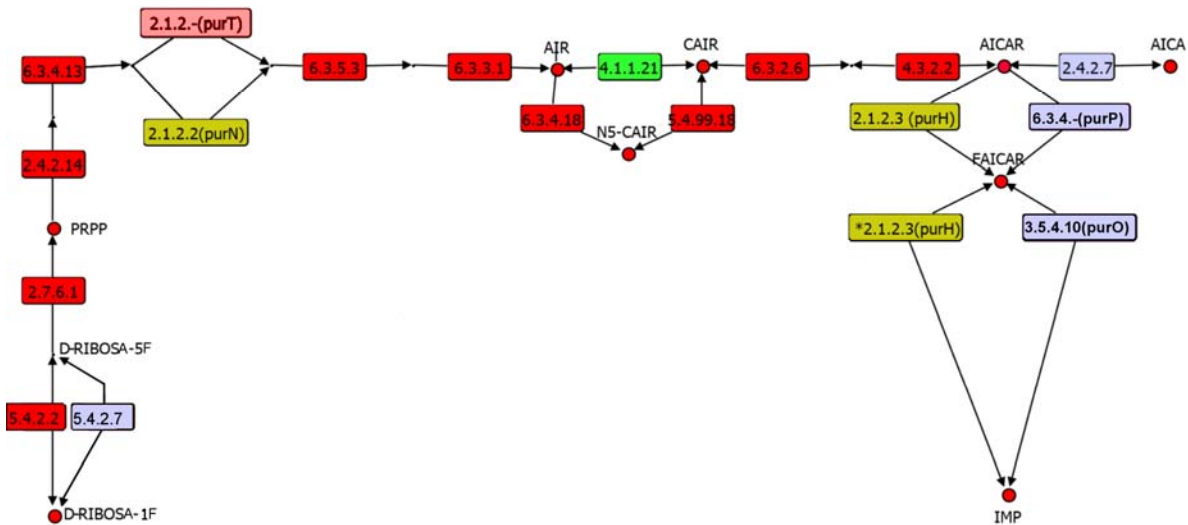


Figura.4- Ruta *de novo* de biosíntesis de purina hacia IMP. En cuadros rojos se observan los pasos enzimáticos asociados al LCA. En rojo claro se observan los pasos enzimáticos donde no podemos definir su presencia o ausencia en el LCA. En color oro se observan los pasos enzimáticos folato-dependientes. En azul los pasos enzimáticos no asociados al LCA de acuerdo a nuestros datos de DT. Etiquetados como purP y purO los pasos enzimáticos específicos de Arqueas. En verde paso enzimático específico de Eucariontes.

Es probable que la familia enzimática PurT estuviera presente en la LCA, con eventos de pérdida en Eucaria debido a su dependencia de formato como sustrato. En este sentido, el formato se describe como un precursor de los primeros donadores de grupos metilo y una de las principales moléculas presentes en las condiciones de prebióticos, antes de metabolismo de folatos³⁸⁻⁴⁰. En una fase posterior, la aparición de la biosíntesis de folato pudo haber facilitado la aparición de PurN (Figura 4, rectángulo color oro), consiguiendo de esta manera la co-ocurrencia de PurN y PurT en el LCA. Trabajos previos han sugerido que PurT precedió a la aparición de PurN, ya que PurT utiliza un sustrato más primitivo que la vía folato dependiente⁴⁰. Adicionalmente, se ha demostrado que el sustrato de PurT,

el formato, es tóxico y se une a la citocromo c-oxidasa en eucariotes⁴¹⁻⁴³, desacoplando las reacciones redox, y con ello favoreciendo la selección de PurN en organismos eucariotes.

Otro paso enzimático identificado en la vía de biosíntesis de purina se consigue mediante la fosforibosilformylglicinamidine sintasa (6.3.5.3) que transforma FGAR a formilglicinamidina-ribonucleótido (FGAM). Esta enzima se compone de dos subunidades catalíticas, PurQ y PurL. En general, esta enzima puede estar presente como una proteína multidominio o como un complejo multiprotéico, donde las subunidades L y Q son codificadas de forma independiente. Nuestro análisis identificó la co-ocurrencia de estas subunidades en los tres dominios celulares. Finalmente, el último paso hacia la transformación AIR se lleva a cabo por la fosforibosilformilglicinamidina ciclo-ligasa (EC 6.3.3.1), que transforma FGAM en AIR, la cual también se encontró estar ampliamente distribuido en los tres dominios celulares, sugiriendo su presencia en el LCA.

El producto de los pasos descritos anteriormente, el intermediario AIR (5-amino-1-5-fosfo-D-ribosil imidazol-4-carboxilato de etilo), puede ser transformado por dos vías; en la primera, una sola reacción, catalizada por la familia enzimática (AIR carboxilasa) asociada al EC 4.1.1.21 (Figura 4, cuadro verde), está involucrada en la síntesis de CAIR; mientras que en la segunda vía es requerido, un camino de dos pasos, asociados a las familias enzimáticas N5-carboxyaminoimidazole ribonucleótido sintasa (EC 6.3.4.18) y 5-(carboxiamino) ribonucleótido imidazol mutasa (EC 5.4.99.18). Cuando se evaluó la distribución taxonómica de estas enzimas, se encontró a la ribonucleótido N5-carboxyaminoimidazole sintasa (6.3.4.18) ampliamente distribuida en los tres dominios celulares, mientras que la familia enzimática 5-(carboxiamino) imidazol ribonucleótido

mutasa (5.4.99.18) presentó una amplia distribución taxonómica en bacterias y arqueas y escasamente en eucariotes, lo que sugiere su presencia en el LCA con su posterior pérdida en Eucaria (Figura 4).

De manera adicional, se ha evidenciado que a nivel secuencia la familia enzimática 5-(carboxiamino) ribonucleótido imidazol (5.4.99.18) mutasa mantiene niveles considerables de similitud con la AIR carboxilasa (4.1.1.21). De hecho, nuestro análisis a nivel de superfamilia estructural muestra que ambas familias enzimáticas comparten un origen evolutivo común, ya que ambas pertenecen a la superfamilia N5-CAIR mutasa (carboxilasa fosforribosil-aminoimidazole). La observación de que la AIR carboxilasa (4.1.1.21) se distribuye exclusivamente en Eucaria sugiere su subfuncionalización a partir de la 5-(carboxiamino) imidazol ribonucleótido mutasa (5.4.99.18) mediante un proceso de divergencia.

El descubrimiento de la aparición de estas ramas es interesante, ya que nos permite inferir procesos asociados a los cambios ambientales. En este contexto, Tribunskikh et al.⁴⁴ contrario a nuestra propuesta, sugiere un orden de divergencia de la familia enzimática fosforribosil-aminoimidazole carboxilasa (4.1.1.21) hacia la 5 - (carboxiamino) imidazol ribonucleótido mutasa (5.4.99.18) como consecuencia de la disminución de CO₂ atmosférico, en conjunto con la adición de la familia enzimática 5-(carboxiamino) ribonucleótido imidazol sintasa asociada al EC 6.3.4.18 y por consecuencia, en la vía de dos pasos. Esos autores fundamentan su propuesta en el hecho de que las vías de dos pasos pueden trabajar a niveles bajos de CO₂, en condiciones aerobias o anaerobias. Sin embargo, nuestros datos sugieren un escenario alternativo, donde, a pesar de que las concentraciones

de CO₂ en la atmósfera primitiva eran altas, las condiciones ácidas en los océanos primitivos favorecerían la formación de HCO₃⁻. Esto es consistente con experimentos que sugieren que la concentración de HCO₃⁻ oscilaba en niveles de 30 a 30.000 veces mayor en los océanos primitivos que en los actuales⁴⁵. Curiosamente, la enzima 6.3.4.18, asociada a la vía de dos pasos, utiliza HCO₃⁻ como sustrato, la cual se considera la forma dominante de CO₂ en los océanos primitivos^{45,46}, lo anterior en conjunto con nuestros datos de distribución taxonómica apoya la noción de que esta vía precedió a la ruta en la que la reacción enzimática catalizada por la vía 4.1.1.21 utilizaba directamente CO₂ como el sustrato. De hecho, se ha demostrado experimentalmente que en soluciones acuosas con altas concentraciones de KHCO₃, AIR se convierte fácilmente en CAIR en ausencia de enzimas^{47,48}. Este mecanismo tiene lugar a través de la acumulación del intermedio N-CAIR (N5-carboxi-AIR o ribonucleótido N5-carboxyaminoimidazole), el cual posteriormente sufre un reordenamiento a CAIR de manera espontánea, Figura 5.

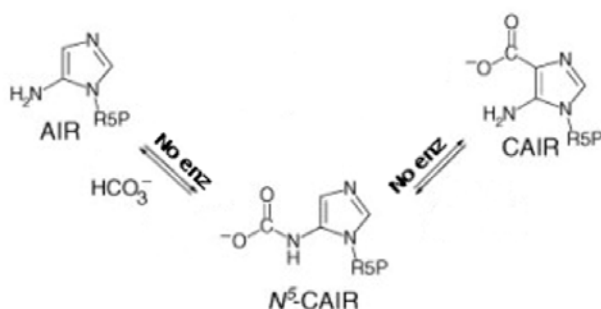


Figura.5- Síntesis química de AIR a CAIR. Las flechas señalan los pasos no enzimáticos que de manera reversible se interconvierten en el intermediaria N-CAIR para su posterior transformación a CAIR. El primer paso de AIR a N-CAIR requiere del exceso del ion bicarbonato para favorecer la síntesis de CAIR.

Estas reacciones parecen ser el molde mediante el cual las actividades enzimáticas se adaptaron formando la vía de dos pasos ampliamente distribuida en las arqueas y las bacterias y probablemente en el LCA. Probablemente, los niveles de HCO_3^- tuvieron una disminución constante debido a la reducción de CO_2 en la atmósfera, como se documenta en la evolución de la atmósfera terrestre⁴⁶. Esta reducción pudo haber seleccionado las enzimas 6.3.4.18 y 5.4.99.18, que transforman AIR en N-CAIR y N-CAIR a CAIR, respectivamente de manera similar a las enzimas madres propuestas en la teoría semi enzimática Lazcano-Miller² de la evolución del metabolismo, Figura 6.

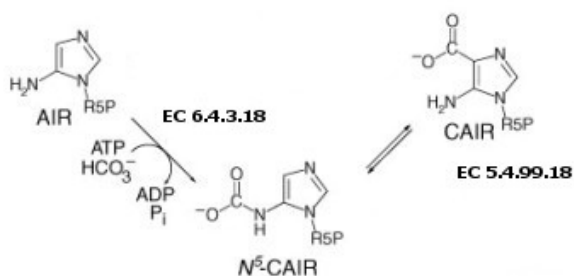


Figura.6- Biosíntesis enzimática de AIR a CAIR asociada al LCA.

Posteriormente, una vez que la atmósfera fue provista de oxígeno, la aparición de las mitocondrias y las células eucariotas fue posible. En este sentido, el HCO_3^- es uno de los principales productos de la respiración mitocondrial, que sigue una conversión pH dependiente a CO_2 ; es decir, la conversión de un anión impermeable en un gas que puede difundir a través de membranas⁴⁹. La transformación de HCO_3^- acumulado en células con actividad mitocondrial a CO_2 pudo haber dado lugar a una regulación más eficiente de pH intracelular en células eucariotas, y en paralelo, al uso de CO_2 como fuente de carbono. En

conjunto, los datos de distribución taxonómica, síntesis química y simulaciones en los océanos primitivos, apoyan la divergencia de la enzima 5.4.99.18 a 4.1.1.21 mediante la selección de un sitio de unión a CO₂, Figura 7.

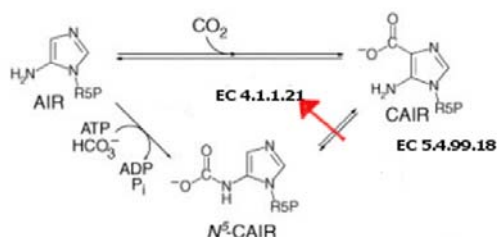


Figura.7- Biosíntesis de AIR a CAIR. La flecha roja indica un probable evento de duplicación de la enzima asociada a la vía de dos pasos (PurE, EC 5.4.99.18) hacia PurE Clase II, dando origen a la vía de un paso específica del dominio eucariote (EC 4.1.1.21).

Los dos últimos pasos en la síntesis de IMP corresponden a la transformación de los intermediarios ribonucleótido 5 - aminoimidazol - 4 - carboxamida (AICAR) y ribotido 5-formamidoimidazole-4-carboxamida (FAICAR), partiendo de N-succinil-5-aminoimidazol-4-ribotido carboxamida (SAICAR) como sustrato. En el primer paso (2.1.2.3), se requiere la actividad AICAR formiltransferasa, mientras que para el segundo paso (3.5.4.10), se requiere la actividad IMP ciclohidrolasa. La enzima bifuncional folato dependiente PurH, se distribuye ampliamente en bacterias y eucaria y parcialmente en arqueas, codificando ambas actividades. PurH exhibe dos sitios catalíticos independientes, en el que cada mitad de la enzima cataliza cada reacción de manera independiente. La transformación a IMP por PurH pudo haber ocurrido posterior a la aparición de la vía de síntesis de folatos (Figura 4, rectángulo color oro). Además, ambas reacciones llevadas a cabo por PurH pudieron haber

surgido en arqueas por diferentes proteínas que catalizan cada paso de forma independiente. En este sentido, la primera reacción se lleva a cabo por la enzima formato dependiente PurP, mientras que la segunda se logra por la IMP ciclohidrolasa PurO (Figura 4, magenta). En ambos casos, el mecanismo catalítico es muy similar a la de su contraparte PurH, y se distribuyen exclusivamente en arqueas; sin embargo, estas dos enzimas no mantienen relaciones en nivel de secuencia o estructural con PurH²³. En resumen, nuestros resultados sugieren que PurP y PurO surgieron en Arqueas, posterior a la divergencia del LCA.

Alternativamente, el cenancestro de las arqueas pudo haber contenido la formato dependiente (PurT) y las familias folato dependientes (PurN-PurH), como se puede observar en las halobacterias. A su vez, PurP pudo haber sido el resultado de un proceso de duplicación de las familias asociadas a la LCA [carbamoil-fosfato sintasa, 6.3.5.5; - fosforribosilamina glicina ligasa, 6.3.4.13; 5 - (carboxiamino) imidazol sintasa ribonucleótido, 6.3.4.18; formiltransferasa fosforibosilglicinamida, 2.1.2.2] las cuales se encuentran presentes en el metabolismo de nucleótidos y poseen homología a nivel estructural con PurP, sustituyendo parcialmente a PurH en algunos clados de arqueas. Por ejemplo, PurH fue identificado en los géneros de *Halobacterias*, *Thermoplasmata* y *Methanomicrobia*, entre otros, incluyendo su co-ocurrencia con PurP, lo que sugiere que sus actividades podrían ser complementarias en estos organismos.

En el caso de PurO, no encontramos relaciones de homología con alguna otra familia enzimática asociada a la biosíntesis de nucleótidos. De manera adicional, se encontró una correlación entre la presencia de PurP y PurO; sin embargo, en algunos clados de arqueas no presentan PurH ni PurO, lo que sugiere que enzimas alternativas podrían realizar la función de PurO y aún permanecen por ser descritas.

9.2.2- Vía de salvamento de Purinas

Vía de salvamento de ribonucleótido de adenina

En la vía de salvamento de purinas, se identificaron diversas rutas para la generación de ribonucleótidos y desoxirribonucleótidos de adenina y guanidina. En este sentido, la primera vía se inicia partiendo de hipoxantina, a través de la hipoxantina-guanina fosforribosiltransferasa (HPRTase) (EC 2.4.2.8) (Figura 8), una enzima con amplia especificidad que se encuentra universalmente distribuida en los tres dominios celulares. Esta enzima logra la transformación bioquímica hacia IMP, partiendo de hipoxantina y PRPP, para la posterior transformación del ribonucleótido de adenina a través de la vía de novo adenil-succinato, que incluye a las enzimas ampliamente distribuidas en los tres dominios celulares (6.3.4.4, la adenilosuccinato-sintetasa; 4.3.2.2, adenilosuccinato liasa; 2.7.4.3, adenilato cinasa; 2.7.4.6, nucleósido difosfato cinasa; 2.7.1.40, piruvato cinasa), sugiriendo su presencia en el LCA.

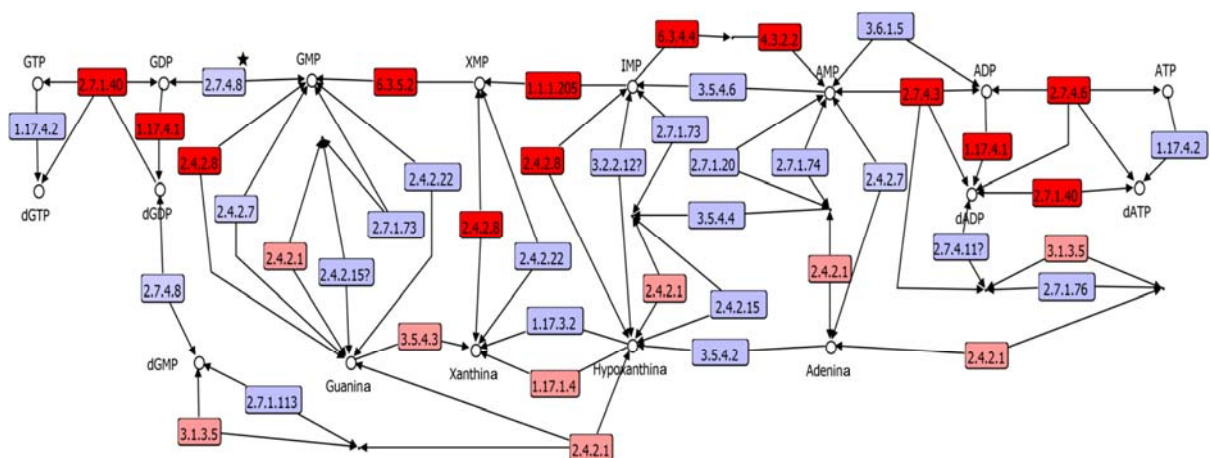


Figura.8- Ruta de salvamento de purinas. En cuadros rojos se observan los números enzimáticos asociados al LCA. Los cuadros en azul representan aquellos números EC no asociados al LCA de acuerdo con nuestros

datos de distribución taxonómica. En color rosa los EC que no se pudieron determinar cómo presentes o ausentes en el LCA. La estrella denota el paso enzimático asociado a la familia enzimática guanilato cinasa.

Para la transformación de alfa-desoxirribonucleótido, hay dos rutas, una asociada con la subunidad beta de la ribonucleósido-difosfato reductasa (subunidad pequeña) (1.17.4.1), en conjunto con la subunidad alfa ribonucleósido-difosfato reductasa (subunidad grande), y la vía 1.17.4.2, que contiene-ribonucleósido trifosfato reductasa adenosilcobalamina-dependiente, y el complejo anaeróbico-ribonucleósido trifosfato reductasa (NrdD y NrdG). La primera de estas vías, para la formación de desoxirribonucleótidos, se encuentra ampliamente distribuida en los tres dominios celulares, lo que sugiere su presencia en el LCA. Por último, las enzimas de la vía 1.17.4.2 se encuentran parcialmente distribuidas en bacterias y arqueas, pero no en Eucaria, lo que hace que sea difícil determinar su presencia en el LCA. Sin embargo, las enzimas asociadas a ambos EC pudieron haber tenido un origen evolutivo en común ya que ambas poseen por lo menos entre algunas de sus subunidades, relaciones de homología estructural.

También es posible sintetizar ribonucleótidos de adenina por una ruta alternativa. Esta vía conserva los tres últimos pasos catalíticos (2.7.4.3, 2.7.4.6 y 2.7.1.40) de la vía descrita anteriormente, excepto el primer paso, el cual es catalizado por la familia adenina fosforribosiltransferasa (PRTase) (2.4.2.7) (Figura 8), esta ampliamente distribuida en bacterias y eucariotes, partiendo de la vía de salvamento de adenina. Esta enzima puede

haberse especializado por el sustrato adenina, posterior a la divergencia del LCA, de un ancestro de amplia especificidad, similar a la HPRTasa (2.4.2.8).

Ruta de salvamento de ribonucleótidos de guanina

A partir de la primera ruta de salvamento, descrita anteriormente, que se inicia con hipoxantina-guanina fosforribosiltransferasa (2.4.2.8), también es posible sintetizar ribonucleótido guanidina, añadiendo a las familias enzimáticas ampliamente distribuidas en los tres dominios celulares: inosina - 5' - monofosfato deshidrogenasa 1 (1.1.1.205) y GMP (glutamina - hidrolizando) sintasa la subunidad A y GMP sintasa (glutamina - hidrolizando) subunidad B (6.3.5.2) (Figura 2). Debido a que la hipoxantina - guanina fosforribosiltransferasa (2.4.2.8) también exhibe especificidad para guanina, es posible sintetizar (guanina monofosfato de ribonucleótido) guanosina 5' - monofosfato (GMP) en un solo paso, a partir de la ruta de salvamento de guanina y PRPP. Sin embargo, aunque el paso enzimático subsecuente al 2.7.4.8, realizado por la familia guanilato cinasa, para la transformación de GMP en GDP se encuentra ampliamente distribuido en eucariotes y bacterias, en arqueas sólo se ha detectado la reacción, pero no la enzima asociada⁵⁰.

Estos datos limitan la posibilidad de hacer comparaciones genómicas y evolutivas con bacterias y eucarias para garantizar la posible presencia de ribonucleótidos de guanina en el LCA (Figura 8, asterisco). Sin embargo, los pasos subsecuentes para la síntesis del GDP a GTP se llevan a cabo con la nucleósido difosfato cinasa (2.7.4.6) y piruvato cinasa (2.7.1.40); estas enzimas se encuentran ampliamente distribuidas en los tres dominios celulares y en conjunto con el paso para la conversión de GTP a dGTP (1.17.4.1) asociado

a la familia enzimática ribonucleósido difosfato reductasa, sugieren la presencia de GTP y dGTP en el LCA.

Por lo tanto, aunque la secuencia del gen asociado con la función de guanilato cinasa en Arqueas aún no ha sido identificado, se sugiere que esta proteína podría haber presentado un origen común con su contraparte en bacterias y eucariotas, ya que la mayoría de las nucleótido cinasas pertenecen a la familia P-loop-hidrolasas. Esta familia muestra una amplia especificidad en el reconocimiento de ribonucleótidos y desoxirribonucleótidos⁵¹

Figura 9.

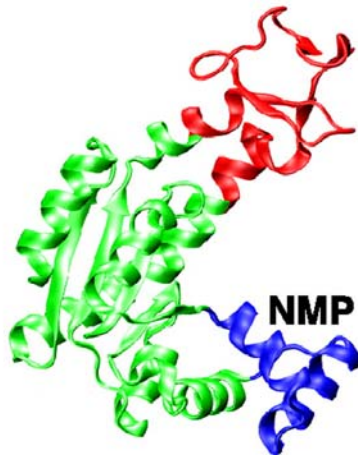


Figura.9- Plegamiento de la familia P-loop. En azul se muestra el dominio de unión a nucleótido monofosfato, el cual se ha observado ser de amplia especificidad. En rojo se observa la tapa y en verde el núcleo de la enzima.

En este sentido, se ha evidenciado que sólo dos mutaciones son suficientes para introducir la actividad adenilato cinasa en una enzima con actividad de guanilato cinasa⁵², lo que sugiere un enmascaramiento de la asignación de la enzima guanilato cinasa por alguna adenilato cinasa. Así, es posible que una enzima con amplia especificidad similar a la

adenilato cinasa (Figura 10, asteriscos) reconociera NMPs para su transformación en NDPs; por lo tanto, la especialización hacia la función de guanilato cinasa pudo haber ocurrido partiendo de una adenilato cinasa de amplia especificidad.

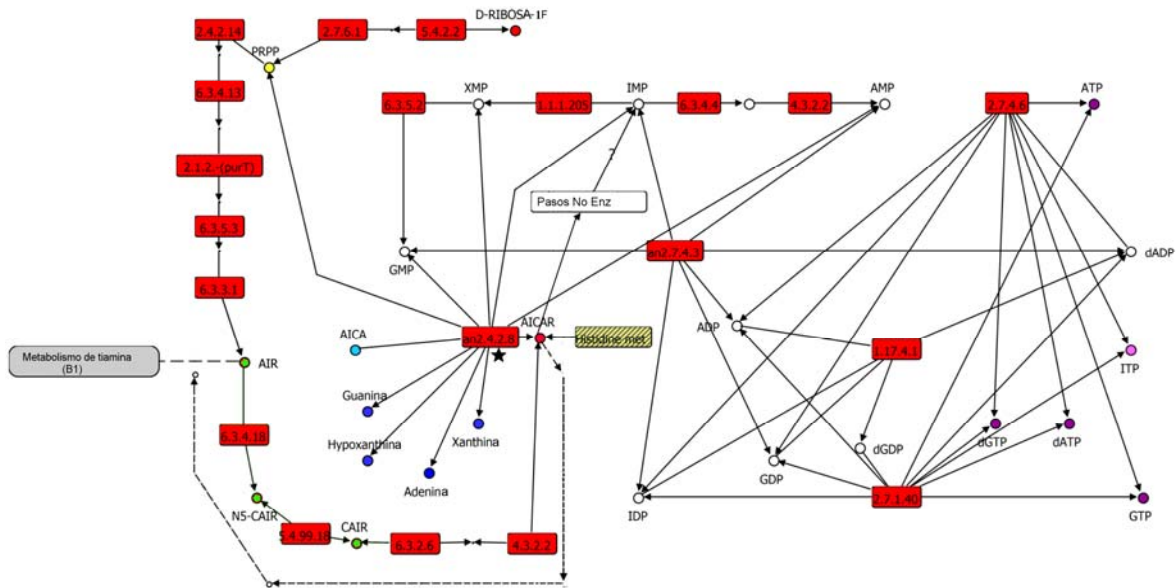


Figura.10- Integración de las vías de novo y salvamento de purinas en el LCA. En cuadros rojos los pasos enzimáticos asociados al LCA. En cuadro blanco la posible ruta semi enzimática presente en etapas tempranas de la evolución celular propuesta por Becerra et al. El cuadro con barras representa la biosíntesis de histidina previamente propuesta como ancestral proporcionado AICAR. En círculos azules los substratos propuesto con origen prebiótico. El metabolismo de tiamina (vitamina B1) se muestra en cuadro gris. En círculos púrpura los ribonucleótidos posiblemente presentes en el LCA. La línea discontinua muestra a AICAR a contraflujo como sustrato semilla para la síntesis de AIR. En estrella se observa el EC 2.4.2.8 asociado a un posible familia ancestral adenilato cinasa de amplia especificidad por sustratos.

En la actualidad, miembros de la familia de adenilato cinasa, en particular, de la subfamilia AK6 se han descrito como proteínas que exhiben la capacidad de transformar diferentes NMPs a sus correspondientes NDPs; específicamente, la subfamilia AK6 transforma AMP,

dAMP, CMP, dCMP, IMP, y GMP a sus correspondientes NDPs y dNDPs^{53,54}. Esta observación también sugiere que una enzima ancestral con amplia especificidad para la síntesis de ribonucleótidos y desoxinucleótidos pudo ser utilizado en el LCA. También es interesante que esta subfamilia pueda utilizar IMP como sustrato, ya que este nucleótido pudo haber tenido un papel activo en la evolución de los ácidos nucleicos y el código genético⁵⁵. Nuestros resultados muestran que en arqueas se detectan secuencias asociadas a la subfamilia tipo AK6, lo que sugiere que la función de guanilato cinasa en estos organismos, podría estar llevándose a cabo al menos en parcialmente por las enzimas de esta subfamilia.

9.2.3- Integrando el análisis de las rutas de salvamento y *de novo* de purinas.

La distribución taxonómica de las enzimas asociadas al metabolismo de purinas muestra que los segmentos de las vías de novo y de salvamento fueron complementarias para disponibilidad de ácidos nucleicos en el LCA antes de la divergencia de los tres dominios celulares. Este resultado se correlaciona con la síntesis química bajo condiciones prebióticas de precursores de bases nitrogenadas para la vía de salvamento de purina, como se describe anteriormente^{22,25,56,57}. (Figura 3, los círculos azules). Estudios previos concuerdan con el hecho de que estas rutas eran dependientes de la biosíntesis de PRPP, ya que se ha argumentado que los nucleósidos y nucleótidos de PRPP, son susceptibles a hidrólisis y, por lo tanto, se consideran muy improbables como compuestos prebióticos. Además, nuestros datos muestran una amplia distribución taxonómica en los tres dominios celulares de las enzimas asociadas con los dos últimos pasos enzimáticos (5.4.2.2 y 2.7.6.1)

necesarias para la transformación de la ribosa-1-fosfato a PRPP (Figura 3, círculo amarillo), lo que apoya la biosíntesis de esta molécula en el LCA.

A pesar de la posible ausencia de folatos y las enzimas folato dependientes PurH y PurN en etapas tempranas de la aparición de vida (Figura 4, cuadros color oro), la mayoría de las enzimas asociadas con la vía *de novo* de la biosíntesis de IMP pueden haber ocurrido en el LCA (Figura 4), con base en su amplia distribución taxonómica en los tres dominios celulares. Posterior a las primeras etapas de la emergencia la vida y la biosíntesis de folatos, la aparición de PurH pudo haber consumado la ruta actual de biosíntesis *de novo* de purinas, completando la pasos enzimáticos entre AICAR e IMP. (Figura 4, PurH se muestra en las cajas de oro). También, el hecho de que PurP y PurO se han identificado exclusivamente en Arqueas, sin homología evidente con PurH, sugiere que estas proteínas se generaron posterior a la divergencia de las arqueas (Figura 4, cuadro Magenta).

Previo a la aparición de la síntesis de folatos y por consecuencia de PurH, es posible que, además de las rutas de salvamento, existieran rutas semi enzimáticas alternativas, a partir de los sustratos 5-amino-4-imidazol carboxiamida (AICA) y PRPP, generando AICAR por medio de una fosforribosiltransferasa ancestral (PRTase), con posteriores transformaciones no enzimáticas a IMP, como se ha propuesto anteriormente²² (Figura 10, las casillas en blanco). En concreto, la adenina fosforribosiltransferasa (PRTases) (2.4.2.7) sintetiza ribonucleótidos de adenina por medio de la vía de salvamento de adenina y también puede transformar AICA a AICAR. Esta enzima pudo haberse especializado posterior a la divergencia del LCA, de un ancestro de amplia especificidad, similar a la HGPRTasa (PRTases) (2.4.2.8); de hecho, esta enzima se encontró ampliamente distribuida en los tres dominios celulares y exhibe un mecanismo catalítico similar. En términos evolutivos, las

enzimas 2.4.2.7 y 2.4.2.8 pertenecen a la superfamilia estructural de fosforribosiltransferasas, lo que sugiere un origen evolutivo común.

Con base a los datos anteriormente descritos, se sugiere que una enzima PRTasa ancestral similar a la HGPRTasa (PRTases) (2.4.2.8) (Figura 10, en estrella) con amplia especificidad por diversos sustratos estructuralmente relacionados, como la guanina, xantina, hipoxantina, adenina, y AICA, estuvo presente en la etapa tempranas de la aparición de la vida. Esta enzima ancestral pudo haber estado involucrada en la transformación del sustrato considerado con orígenes prebióticos AICA (Figura 10, círculo azul claro) en AICAR, no sólo para alimentar la síntesis de IMP de manera semi enzimática, según propuestas anteriores²² (Figura 10, la caja blanca), sino también en mayor medida a la biosíntesis de tiamina (vitamina B1). Es interesante notar que el metabolismo de tiamina pudo haber sido alimentado en dos direcciones: como una bifurcación de la ruta *de novo* hacia IMP y también a contraflujo utilizando AICA como sustrato de la semilla (Figura 10, siguiendo la dirección de las flechas discontinuas). De manera adicional, el metabolismo de tiamina pudo haber sido alimentado desde una tercera fuente por medio del metabolismo de histidina (Figura 10, la cuadro con barras) ; la conexión con la biosíntesis de purina resulta del paso enzimático catalizado por la glicerol fosfato imidazol (IGP) sintasa, que transforma N - (5- fosforribosil) - formimino - 5 - aminoimidazol - 4 - carboxamida ribonucleótido (PRFAR) en AICAR, el cual se recicla en la ruta biosintética *de novo* de purina, y en imidazol - glicerol 3 - fosfato, que conduce a biosíntesis de histidina. De manera interesante, trabajos previos han sugerido también que la ruta de biosíntesis de histidina es antigua y se relaciona a etapas tempranas de la evolución celular, inclusive antes de la aparición del LCA¹⁰.

Nuestra propuesta de una rama ancestral relacionada con la síntesis de tiamina es consistente con las funciones catalíticas que se han propuesto a esta molécula en la evolución temprana de la vida, dado su papel catalítico esencial en la mayoría de los organismos y su indispensabilidad en varios puntos centrales del metabolismo anabólico y catabólico⁵⁸, así como en posibles vías semi enzimáticas que podrían preceder las actuales². Es interesante que la aparición de esta rama en etapas temprana de la vida y antes de la emergencia del LCA pudo haber alimentado el metabolismo de tiamina de manera semi enzimática, ya que la transformación AICAR en AIR puede ocurrir fácilmente por una vía de síntesis química⁵⁹.

En este sentido, las moléculas de tiamina o sus derivados se unen al mARN sin necesidad de cofactores o proteínas en los tres dominios celulares, formando un complejo que secuestra el sitio de unión al ribosoma y que sugiere la existencia de una forma ancestral de *riboswitches*, implicado en mecanismos de regulación⁶⁰. Además, la tiamina pudo haber interactuado con el ARN, dando lugar a ribosimas catalíticamente versátiles relacionadas con el mundo del ARN, debido a sus capacidades de unión de ARN y catalíticas⁵⁸.

Por último, se sugiere que la función de la guanilato cinasa pudo haber sido llevado a cabo en el LCA por una enzima ancestral con amplia especificidad por NMPs, similares a los de la subfamilia AK6 de adenilato cinasa. En paralelo, esta enzima ancestral pudo haber proporcionado nucleótidos de inosina (Figura 10, círculo rosa), al cual se le ha propuesto haber tenido un papel activo en una etapa rudimentaria del código genético⁵⁵. El ITP puede competir o reemplazar los sitios de unión de ATP y GTP en diversas proteínas, tales como la ARN polimerasa^{61,62}, que pueden haber retenido su afinidad y especificidad actual por ITP como remanente de su papel ancestral.

Adicionalmente, la inosina mantiene una fuerte similitud estructural con la guanina⁶³ (Figura suplementaria S3), y puede incluso tener la misma función al apareado de la guanina con la citosina en los codones. Actualmente, la inosina se ha encontrado en la tercera posición de los anticodones, apareando con los codones en las bases T, C o A y disminuyendo de ese modo la necesidad de 61 tARN's uno para cada codón. En este sentido, se ha sugerido que la inosina puede producirse por desaminación de adenosina o incluso en la catálisis mediada por ARN en etapa temprana de la aparición del código genético, y que fue excluido en los ácidos nucleicos cuando el apareado canónico Watson-Crick evolucionó para evitar reglas ambiguas en la replicación⁵⁵. Posteriormente, la especialización y la divergencia hacia guanilato cinasa de miembros con amplia especificidad de una subfamilia ancestral, pudo haber aumentado la disponibilidad de guanina, favoreciendo la sustitución de inosina por guanina.

9.3- Evolución del Metabolismo de Pirimidinas

9.3.1- Biosíntesis *de novo* de pirimidinas

Con base en su distribuciones taxonómicas, se evaluaron las siguientes familias enzimáticas asociadas con la biosíntesis *de novo* de pirimidinas: carbamoil -fosfato sintasa cadena grande, carbamoil-fosfato sintasa cadena pequeña, 6.3.5.5; aspartato carbamoiltransferasa 1, 2.1.3.2; dihidroorotasa, familia DHOase, 3.5.2.3; dihidroorotato deshidrogenasa B (NAD⁺), la subunidad catalítica, dihidroorotato oxidasa B, la transferencia de electrones de la subunidad, 1.3.1.14 ; dihidroorotato deshidrogenasa (quinona), 1.3.5.2 ; dihidroorotato deshidrogenasa (fumarato), 1.3.98.1; orotato fosforribosiltransferasa 1, PyrE1, 2.4.2.10 ; orotidina 5'-fosfato descarboxilasa, 4.1.1.23; quinasa citidilato, 2.7.4.14; nucleósido difosfato quinasa, 2.7.4.6 (Figura 11). Estas enzimas forman la rama de biosíntesis *de novo*

hacia UTP partiendo de L-glutamina. La rama entera pudo haber ocurrido en el LCA, debido a la amplia distribución taxonómica de las enzimas que la componen (Figura 11, cuadros de color rojo). Uno de los pasos más interesantes de esta vía es la conversión de dihidroorotato a orotato, que se lleva a cabo por enzimas clasificadas en dos clases de dihidroorotato deshidrogenasas (DHODs), de acuerdo a su aceptor terminal de electrones y sus relaciones a nivel de secuencia⁶⁴.

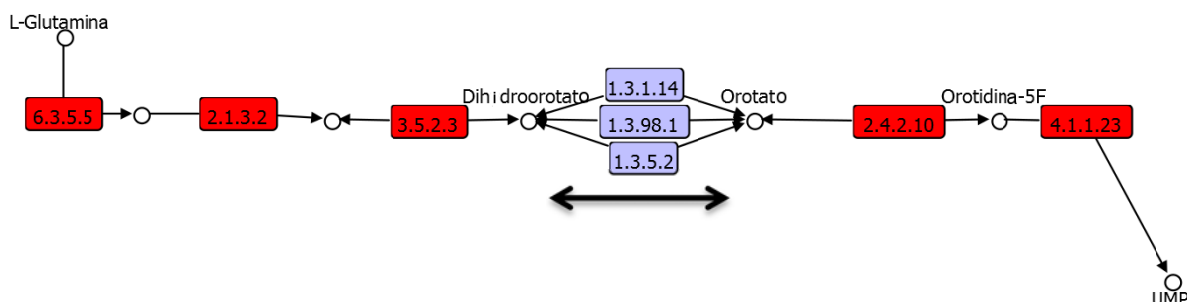


Figura.11- Vía de biosíntesis de pirimidinas. En cuadros rojos se muestran los números enzimáticos asociados al LCA y en azul, subrayados por la flecha, los pasos enzimáticos asociados a las familias de DHODs las cual por datos de comparación de familias estructurales se encontraron ser homólogos y asociado al LCA.

En este sentido, la clase 1 utiliza aceptores de electrones solubles. Esta clase se encuentra ampliamente distribuida en las bacterias gram-positivas, arqueas y en algunos eucariotes unicelulares. A su vez, esta clase se subdivide en la familia DHODA (1.3.98.1) y la familia DHODB (1.3.1.14), que son homodiméricas y heterotetraméricas, respectivamente. La DHODA utiliza fumarato como aceptor de electrones, mientras que DHODB utiliza NAD^+ ⁶⁵. Los miembros de la clase 2 (1.3.5.2) se encuentran anclados a la membrana celular externa y utilizan quinonas de la cadena respiratoria como aceptor de electrones⁶⁶.

Estas enzimas se encuentran principalmente en la mayoría de los organismos eucariotas y bacterias gram-negativas, de acuerdo con nuestros resultados de distribución taxonómica.

A pesar de estas diferencias, ambas familias de DHODs pertenecen a la superfamilia estructural FMN-linked oxidoreductasas, según nuestro análisis de secuencias en la base de datos de familias estructurales Superfamily, lo que sugiere un ancestro en común.

La distribución taxonómica de las familias 1 y 2 correlaciona con su presencia en bacterias con una membrana y de doble membrana, respectivamente. En particular diversos grupos sugieren la transición de organismos bi-membrana a mono-membrana⁶⁷⁻⁶⁹ lo que podría explicar el por qué la familia DHODS fue la única familia enzimática que se vio en la necesidad de generar una clase diferente (la clase 2 es la única asociada a la membrana externa en la síntesis hacia UMP, y la pérdida de la membrana externa pudo ser una presión de selección hacia una nueva clase soluble de DHODS (clase 1)); sin embargo, con nuestros datos de distribución taxonómica no podemos ser concluyentes al respecto. En este sentido la divergencia de esta familia enzimática podría estar asociada a cambios en la membrana celular por lo que sería interesante analizar en detalle la historia evolutiva de esta familia de proteínas, así como los procesos que dieron lugar a su divergencia.

Las DHODA utilizan fumarato como aceptor terminal de electrones, el cual es usado en lugar de oxígeno para la producción de succinato, uno de los procesos esenciales que controlan la *homeostasis* redox para muchos parásitos que viven en condiciones anaeróbicas⁶⁶.

Una amplia distribución taxonómica de la familia enzimática CTP sintasa (6.3.4.2), que transforma UTP en CTP, se encontró en los tres dominios celulares, en conjunto con las

familias enzimáticas para los pasos subsiguientes, la nucleósido difosfato cinasa (2.7.4.6), tiorredoxina NTRA reductasa (1.8.1.9) y la ribonucleótido reductasa (1.17.4.1), dando lugar a desoxi-CTP. Con base en la amplia distribución taxonómica de estas enzimas, se sugiere la presencia de UTP, CTP y desoxi - CTP en el LCA (Figura 12, cajas de color rojo).

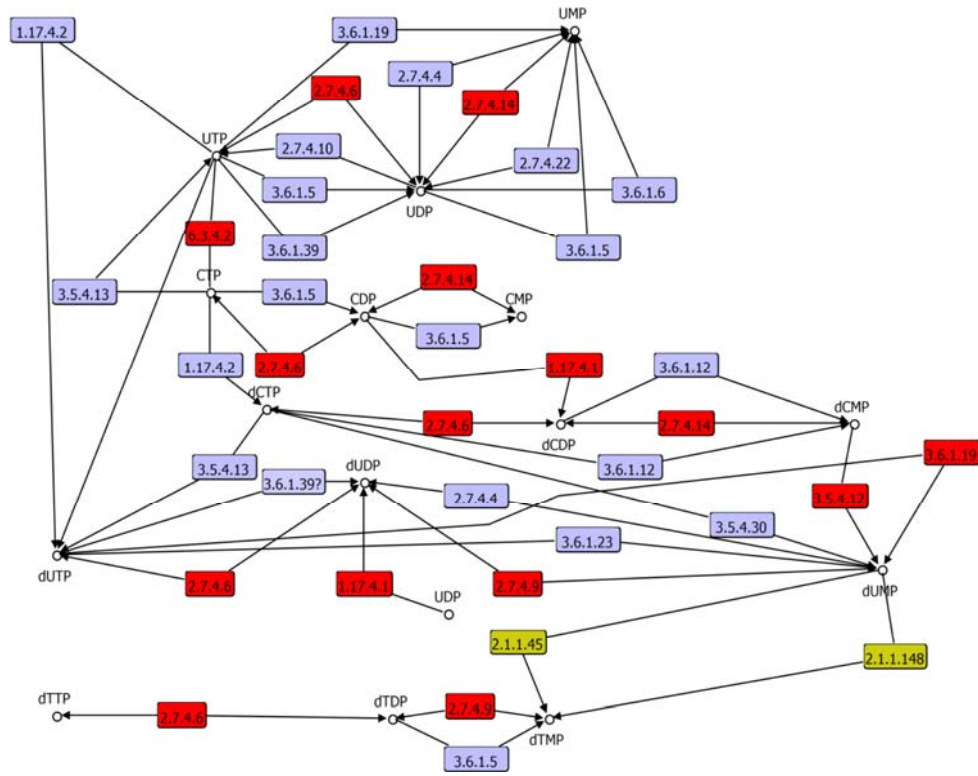


Figura.12- Continuando con la biosíntesis de pirimidinas partiendo de UMP. En rojo pasos enzimáticos asociados al LCA, en azul pasos enzimáticos no asociados al LCA de acuerdo a nuestros datos de distribución taxonómica. En cuadros color oro se representan los pasos enzimáticos asociados a enzimas folato dependientes.

Con respecto a la transformación de desoxiuridina, se han descrito dos rutas principales. Una vía se inicia desde desoxi-CDP, mientras que la segunda inicia de desoxi-CMP

pudiéndose extender a la transformación de desoxiuridina. La ruta de desoxi-CDP necesita dos pasos enzimáticos: el primero requiere de la difosfato de nucleósido cinasa (2.7.4.6), que produce dCTP, y finalmente, este es transformado a dUTP por la desaminasa trifosfato de desoxicitidina (3.5.4.13), esta familia enzimática se encuentra ausente en eucariotas y parcialmente distribuido en bacterias y arqueas. Es difícil determinar la presencia de esta vía hacia la transformación de desoxiuridina en el LCA, debido a la compleja historia evolutiva de la familia enzimática asociada al EC (3.5.4.13), como comentamos anteriormente, Figura 12.

La segunda vía convierte dCMP a dUMP a través de la familia enzimática citidina-desaminasa / desoxicitidilato, que se asocia al EC 3.5.4.12, y cuyos miembros se encuentran ampliamente distribuidos en los tres dominios celulares sugiriendo la posible presencia de esta vía en el LCA, Figura 12.

Por último, la transformación de la desoxi-UMP a desoxi-TMP puede llevarse a cabo a través de dos enzimas folato dependientes, la timidilato sintasa ThyX (2.1.1.148) y timidilato sintasa ThyA (2.1.1.45)²³ (Figura 12 en cuadros color oro). Estas enzimas no son homólogas, lo que sugiere un origen evolutivo independiente. Ambas enzimas promueven la metilación mediante el uso de la 5,10- metilentetrahidrofolato (CH₂-H₄ folato) como donador de carbono. ThyA también utiliza CH₂-H₄ folato para producir dihidrofolato (H₂-folato). Por el contrario, ThyX utiliza el dinucleótido flavina adenina (FAD) y NAD (P) H como cofactores para producir tetrahidrofolato reducido (H₄- folato)⁷⁰. ThyX se distribuye parcialmente en bacterias y arqueas, pero está ausente en eucariotas; su contraparte, ThyA, se distribuye parcialmente en bacterias, escasamente en arqueas y ampliamente en eucariontes. En este contexto, se ha sugerido que las diferencias catalíticas entre ThyA y

ThyX han influenciado la evolución de los genomas de bacterias ^{70,71}. En este sentido, ThyA llega a ser 10 veces más eficiente catalíticamente que ThyX, y estudios previos en más de 400 genomas procariontes han revelado que la capacidad catalítica asociada con ThyA se correlaciona con su presencia en los organismos con tamaños de genoma grande ^{70,71}.

Nuestros resultados de distribución taxonómica son consistentes con los de estudios anteriores, que muestran un patrón de anticorrelación para la presencia / ausencia de estas dos familias de enzimas a través de clados bacterianos, donde también se han observado eventos de reemplazo mutuo ⁷², lo que indica que el metabolismo bacteriano ha modulado el tamaño y la composición de los genomas bacterianos. Además, nuestros datos muestran la ausencia completa de ThyX y la presencia absoluta de ThyA en genomas eucariotas, lo que sugiere que la influencia de ThyA ha sido un factor determinante en el tamaño del genoma eucariota y en el potencial evolutivo de estos organismos. En este contexto, es difícil determinar si algunas de estas familias o ambas estuvieron presentes en el LCA, en función de su compleja distribución taxonómica observada.

9.3.2- Rutas de salvamento Pirimidinas

En la vía de salvamento del ribonucleótido de uracilo existen dos rutas principales que parten del salvamento de uracilo. La primera vía comprende los pasos asociados a la uridin cinasa (2.7.1.48) y la uridin fosforilasa (2.4.2.3). La uridin fosforilasa se encontró escasamente distribuido en los tres dominios celulares y utiliza la base de uracilo como sustrato. La uridin cinasa está ampliamente distribuida en bacterias y eucariotes y escasamente distribuida en arqueas. Por lo tanto, basados en los patrones de distribución taxonómicos, es difícil determinar si todas estas enzimas estaban presentes en el LCA,

debido a su baja distribución en los genomas de Arqueas, que también puede ser un reflejo de eventos de transferencia horizontal.

La familia enzimática de uridin fosforilasa utiliza uridina como un sustrato, que puede ser proporcionada por otra rama vía asociada con la citidina-desaminasa (3.5.4.5) y -pirimidina nucleósido fosforilasa (2.4.2.2). Esta rama utiliza citosina como sustrato de inicio. La citidina deaminasa (3.5.4.5) mantiene una distribución taxonómica similar a la uridina cinasa, mientras que la pirimidina nucleósido fosforilasa (2.4.2.2) sólo se encuentra parcialmente distribuida en bacterias, lo que sugiere que esta vía no estaba presente en el LCA. Existe una tercera vía de salvamento, donde un solo paso enzimático, asociado a la familia uracilo fosforribosiltransferasa (2.4.2.9), es requerido. Esta enzima se encuentra ampliamente distribuida en los tres dominios celulares, lo que sugiere su presencia en el LCA (Figura 13B).

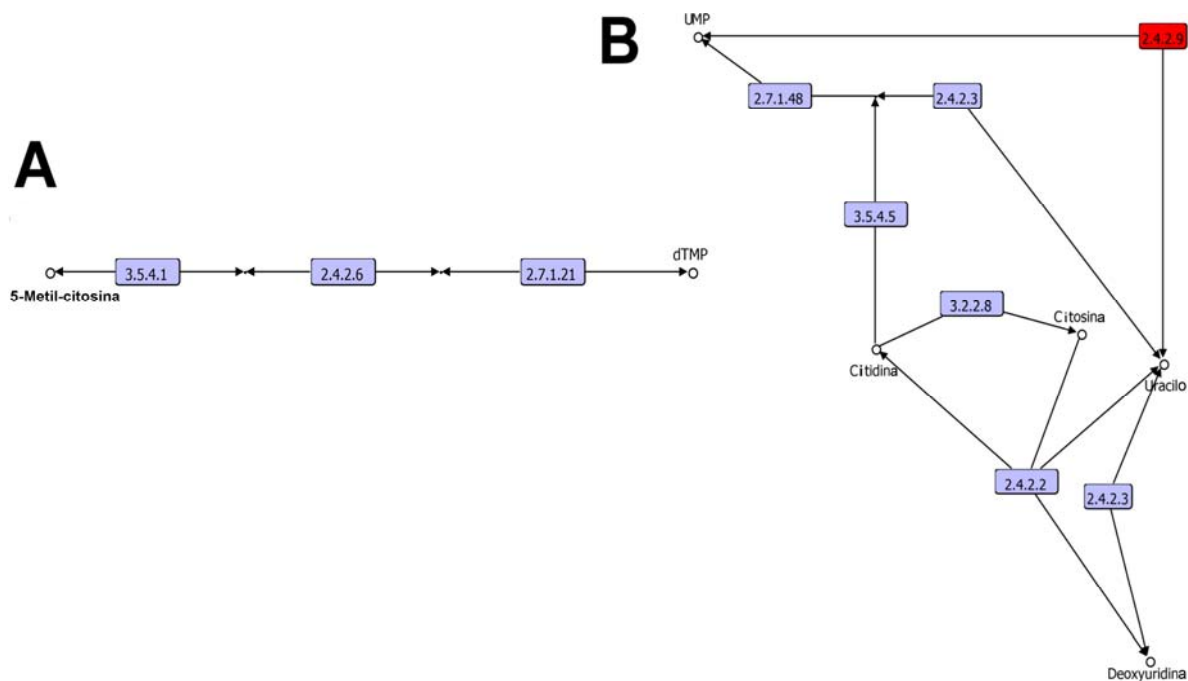


Figura.13- Rutas de salvamento de pirimidinas. En la parte A) de la figura se observan las ruta de salvamente hacia timina. En la parte B) se observan las rutas de salvamento hacia uracilo. En cuadros rojos se indican los pasos enzimáticos asociados al LCA. En azul pasos enzimáticos no asociados al LCA, de acuerdo a nuestros datos de distribución taxonómica.

En cuanto a las rutas de salvamento de ribonucleótidos de citosina, la vía principal se produce en dos pasos, que implican a la uridina - citidina cinasa (2.7.1.48) y - pirimidina nucleósido fosforilasa (2.4.2.2), partiendo de citosina como sustrato. Con base en su distribución taxonómica, la primera familia enzimática, la uridina citidina cinasa, como se ha descrito anteriormente no presenta un patrón de distribución claro para ser considerada en el LCA, mientras que el segundo paso se encuentra sólo escasamente distribuido en bacterias. Además, hay otra vía de salvamento de desoxirribonucleótido de citosina que comienza a partir de la desoxicitidina, en el que la desoxiadenosina / desoxicitidina cinasa (2.7.1.74) participa (Figura 13B), encontrándose sólo escasamente distribuida en Eucaria.

En la vía de salvamento de desoxitimina, dos pasos enzimáticos se encuentran involucrados, el primer paso se encuentra asociado a la familia timidina cinasa (2.7.1.21) y el segundo a la timidina fosforilasa (2.4.2.4), utilizando timina como sustrato semilla (Figura 13A). La enzima asociada con el primer paso se distribuye parcialmente en bacterias y se encuentra escasamente distribuido en eucariotes y arqueas, mientras que la enzima asociada al segundo paso se encuentra escasamente distribuida en los tres dominios celulares. Es difícil determinar si esta vía estuvo presente en el LCA o si su distribución es un reflejo de la transferencia horizontal de genes o pérdida masiva de genes, debido a la

compleja distribución taxonómica de las enzimas pertenecientes a esta vía en los tres dominios celulares.

9.3.3- Integrando el análisis evolutivo de las vías *de novo* y salvamento del metabolismo de pirimidinas.

Con base en las distribuciones taxonómicas de las enzimas asociadas a las vías *de novo* y salvamento de pirimidinas, fue posible identificar, además de la vía *de novo*, una vía de salvamento ampliamente distribuida en los tres dominios celulares (a través de 2.4.2.9) para la síntesis de ribonucleótido de uracilo (Figura 13A), lo que sugiere que el LCA tenía dos vías para llevar a cabo su síntesis. Por otro lado, para la síntesis de desoxirribonucleótidos de uracilo, se encuentran dos rutas principales, de un solo paso; la primera se parte de dCTP, que se transforma por la familia enzimática asociada al EC 3.5.4.13, y la segunda ruta se inicia a partir de dCMP, que es transformado por la enzima asociada al EC 3.5.4.12. La amplia distribución taxonómica en los tres dominios celulares de la familia enzimática asociada al EC 3.5.4.12 sugiere que la segunda vía estaba presente en el LCA (Figura 14).

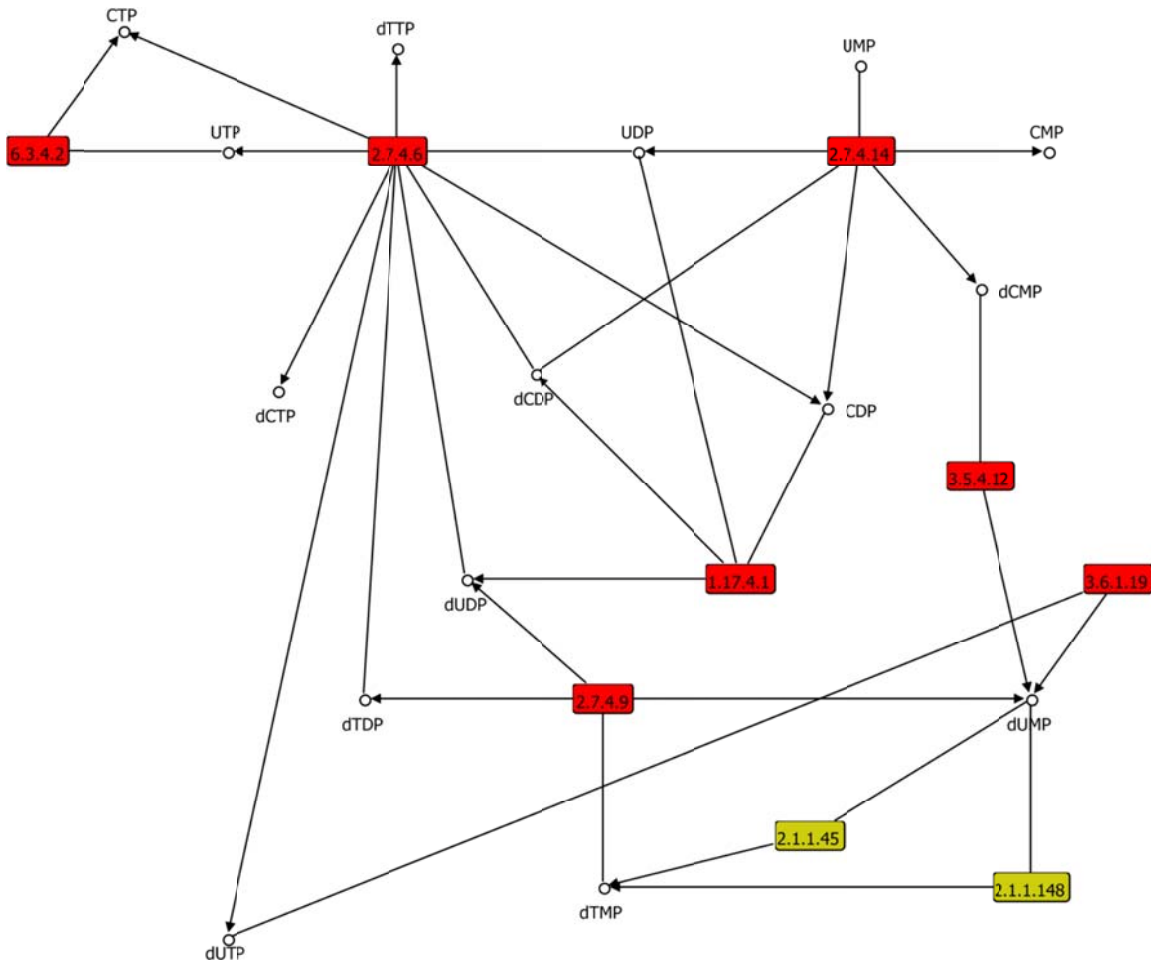


Figura.14- Vía de biosíntesis de pirimidina asociada al LCA. En color rojo se muestran los pasos enzimáticos asociados al LCA. En cuadros color oro enzimas folato dependientes.

Aunque las bacterias gram negativas y eucariotas contienen la subfamilia 2 de DHODS, no es posible deducir una cercana relación evolutiva, ya que los miembros de esta subfamilia en Eucaria se encuentran anclados a la membrana mitocondrial. En este sentido, la adquisición mitocondrial en Eucaria ha sido descrita como un probable evento de transferencia lateral, como se describe en la teoría endosimbiótica⁷³. Por lo tanto, nuestros

resultados apoyan la noción de fagocitosis de una bacteria gram-negativa por una célula protoeucariótica, con su posterior especialización a mitocondrias.

Adicionalmente, la biosíntesis *de novo* de ribonucleótidos y desoxirribonucleótidos se encontró asociada al LCA. En la biosíntesis *de novo* de timina, las enzimas folato dependientes ThyA y ThyX pueden catalizar la transformación de este metabolito de forma independiente, lo que sugiere que esta vía apareció posterior a la aparición de la biosíntesis de folato. Para las enzimas presentes en la vía de salvamento timina, no fue posible determinar su presencia en el LCA.

10- CONCLUSIONES

El análisis presentado se basa en un estudio exhaustivo de genomas completos pertenecientes a organismos de los tres dominios celulares, Bacteria, Arquea y Eucaria, asociado al conocimiento bioquímico acumulado del metabolismo de nucleótidos. Este análisis nos permitió identificar aspectos relacionados con su origen y evolución. Una de nuestras principales conclusiones es que determinamos varias enzimas del metabolismo de nucleótidos como ampliamente distribuidas en los tres dominios celulares reflejando la antigüedad de algunos segmentos de las vías de salvamento y *de novo* de purinas, las cuales pudieron haber sido complementarias y estarían estrechamente asociadas en el LCA. Además, se identificó una gran parte de la rama *de novo* de purinas como ampliamente distribuida en bacterias, arqueas y eucariotas, hacia la biosíntesis de IMP (un precursor clave de purinas). Esta rama pudo haber estado asociada, en las primeras etapas de la evolución de celular, con el metabolismo de la tiamina (vitamina B1) y posteriormente fue

complementado por la adición de dos nuevos pasos enzimáticos para completar la ruta de biosíntesis de IMP por medio de la familia enzimática folato dependiente PurH, dando lugar a la biosíntesis contemporánea *de novo* de purinas.

La ancestría y la divergencia de las enzimas asociadas con estas rutas ofrecen pistas sobre los cambios ambientales en las primeras etapas de la evolución celular. Tal es el caso de la divergencia de la familia enzimática fosfo-ribosil-aminoimidazol carboxilasa (4.1.1.21) a partir de la ribonucleótido N5-carboxiaminoimidazol mutasa (5.4.99.18); el origen y divergencia de estas dos familias enzimáticas apoya la hipótesis del origen de la vida en los mares primitivos con altos niveles de HCO_3^- . Una vez que la atmósfera fue provista de oxígeno y los primeros organismos eucariotas con mitocondrias surgieron, la familia enzimática asociada al EC 5.4.99.18 divergió al EC 4.1.1.21 a través de la adquisición de un sitio de unión de CO_2 .

En el caso de las pirimidinas, fue posible inferir en el LCA la biosíntesis de ribonucleótidos de uracilo, tanto por las vías *de novo* como de salvamento, lo que sugiere que este ribonucleótido pudo haber estado involucrado en un gran número de funciones enzimáticas y / o en la regulación enzimática como vestigio del mundo de ARN. Además, fue posible asociar la síntesis de desoxirribonucleótidos de citosina y uracilo en el LCA, y una vez que la biosíntesis de folatos fue posible, pudo suceder la biosíntesis desoxirribonucleótidos de timina, debido a su dependencia enzimática por folatos. Antes de la aparición de la biosíntesis de folatos, algunas variantes de estas tres bases, incluyendo citosina o derivados de citosina metilados, con base a su similitud con la timina, pudieron haber desempeñado un papel similar al actual de la timina en el ADN. Esta inferencia es apoyada por el hecho de que varios análogos de citosina y uracilo parcialmente se integran en el ADN como

sustitutos de timina⁷⁴⁻⁷⁶. Por lo tanto, la timina al no tener un papel activo en la transcripción de nucleótidos presenta menor compromiso funcional al ser remplazada. Este argumento es consistente con la evolución de cepas bacterianas⁷⁷, donde se demuestra que pueden generarse cepas con la capacidad de incorporar derivados de uracilo (cloro - uracilo) mediante el remplazo de hasta el 98 % de la timina y mantener la viabilidad celular.

En relación con la síntesis de purinas, nuestros resultados revelaron que la biosíntesis de adenina podría haberse llevado cabo en el LCA por la familia adenilato cinasa. En el caso de la biosíntesis de guanina, encontramos una compleja historia evolutiva; por ejemplo, su síntesis se ha detectado en arqueas, pero no la secuencia del gen⁵⁰. El hecho de que no se ha identificado la guanilato cinasa de tipo bacteriana en arqueas sugiere tres posibles escenarios: a) una alta divergencia génica, b) un enmascaramiento de la función por alguna otra familia enzimática, y / o c) una pérdida de gen o un desplazamiento no ortólogo. Como se discutió anteriormente, es probable que la función de guanilato cinasa ha sido enmascarado por la adenilato cinasa de la subfamilia tipo AK6.

Esta enzima es homóloga a la guanilato cinasa canónica de eucaria y bacterias y es probablemente la más cercana a una posible enzima ancestral presente en el LCA, con promiscuidad en la síntesis de los NMP y dNMPs a sus respectivos NDPs y dNDPs. La similitud estructural de la inosina y guanina, junto con la afinidad remanente de inosina en varias proteínas tales como ARN polimerasa, además de nuestros datos de distribución taxonómica que muestran una posible ruta de LCA para la síntesis de ITP (Figura 10, círculo magenta) por medio de una enzima ancestral de amplio especificidad, sugieren que esta base pudo haber jugado un papel importante en la evolución temprana de la célula,

como se ha propuesto anteriormente⁵⁵. La enzima ancestral pudo haber proporcionado tanto IDP y GDP para su transformación posterior a NTPs, jugando un papel importante en el procesamiento de información similares en el código genético debido a sus parentescos estructurales.

La especialización posterior de guanilato cinasa pudo haber generado una mayor disponibilidad de nucleótidos de guanidina facilitando el reemplazo de la inosina, evitando las reglas ambiguas en la replicación del ADN y alcanzándose una etapa de consolidación del código genético en el LCA.

11- PERSPECTIVAS

Este trabajo indica que probablemente el metabolismo de tiamina pudo estar presente en etapas tempranas de la evolución celular, lo que denota la importancia de estudios evolutivos posteriores. También sería interesante evaluar la función de las enzimas identificadas como hits positivos de adenilato cinasas tipo AK6 en arqueas, ya que podrían estar llevando acabo la función de guanilato cinasa en estos organismos. La correlación de las DHDOHS en organismos uni-membrana y bi-membrana sugiere que su divergencia en distintas clases podría estar asociada a transiciones relacionadas con cambios en la membrana plasmática por lo que sería importante estudiar a profundidad los procesos que dieron lugar a las transformaciones de esta familia. Varias familias enzimáticas pertenecientes al metabolismo de nucleótidos utilizan moléculas derivadas del metabolismo de folatos como precursores de reacción, lo que denota la relevancia de estudios evolutivos posteriores para este metabolismo.

12- REFERENCIAS

- 1 Nelson, D. L., Lehninger, A. L. & Cox, M. M. *Lehninger principles of biochemistry*. (Macmillan, 2008).
- 2 Lazcano, A. & Miller, S. L. On the origin of metabolic pathways. *J Mol Evol* **49**, 424-431 (1999).
- 3 Haldane, J. B. The origin of life. *Rationalist Annual* **148**, 3-10 (1929).
- 4 Oparin, A. Proiskhozhedenie Zhizni. *Mosckovskii Rabochii, Moscow. Reprinted and translated in Bernal JD (1967) The origin of life. Weidenfeld and Nicolson, London (1924)*.
- 5 Caetano-Anolles, G. *et al.* The origin and evolution of modern metabolism. *The international journal of biochemistry & cell biology* **41**, 285-297, doi:10.1016/j.biocel.2008.08.022 (2009).
- 6 Horowitz, N. H. On the evolution of biochemical syntheses. *Proceedings of the National Academy of Sciences of the United States of America* **31**, 153 (1945).
- 7 Jensen, R. A. Enzyme recruitment in evolution of new function. *Annual review of microbiology* **30**, 409-425, doi:10.1146/annurev.mi.30.100176.002205 (1976).
- 8 Lazcano, A. & Miller, S. L. The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* **85**, 793-798 (1996).
- 9 Lopez de la Osa, J. *et al.* Getting specificity from simplicity in putative proteins from the prebiotic earth. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14941-14946, doi:10.1073/pnas.0706876104 (2007).
- 10 Fani, R. & Fondi, M. Origin and evolution of metabolic pathways. *Phys Life Rev* **6**, 23-52, doi:DOI 10.1016/j.plrev.2008.12.003 (2009).
- 11 Diaz-Mejia, J. J., Perez-Rueda, E. & Segovia, L. A network perspective on the evolution of metabolism by gene duplication. *Genome biology* **8**, R26, doi:10.1186/gb-2007-8-2-r26 (2007).
- 12 Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature reviews. Genetics* **5**, 101-113, doi:10.1038/nrg1272 (2004).
- 13 Zhou, T. Computational reconstruction of metabolic networks from KEGG. *Methods Mol Biol* **930**, 235-249, doi:10.1007/978-1-62703-059-5_10 (2013).
- 14 Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555, doi:10.1126/science.1073374 (2002).
- 15 Ouzounis, C. A., Kunin, V., Darzentas, N. & Goldovsky, L. A minimal estimate for the gene content of the last universal common ancestor--exobiology from a terrestrial perspective. *Research in microbiology* **157**, 57-68, doi:10.1016/j.resmic.2005.06.015 (2006).
- 16 de Crecy-Lagard, V., El Yacoubi, B., de la Garza, R. D., Noiriél, A. & Hanson, A. D. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *Bmc Genomics* **8**, doi:Artn 245
Doi 10.1186/1471-2164-8-245 (2007).
- 17 Hernandez-Montes, G., Diaz-Mejia, J. J., Perez-Rueda, E. & Segovia, L. The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution. *Genome biology* **9**, R95, doi:10.1186/gb-2008-9-6-r95 (2008).

- 18 Kyrpides, N., Overbeek, R. & Ouzounis, C. Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* **49**, 413-423 (1999).
- 19 Delaye, L., Becerra, A. & Lazcano, A. The last common ancestor: what's in a name? *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **35**, 537-554, doi:10.1007/s11084-005-5760-3 (2005).
- 20 Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. *Nature genetics* **21**, 108-110, doi:10.1038/5052 (1999).
- 21 Fitz-Gibbon, S. T. & House, C. H. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic acids research* **27**, 4218-4222 (1999).
- 22 Becerra, A. & Lazcano, A. The role of gene duplication in the evolution of purine nucleotide salvage pathways. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **28**, 539-553 (1998).
- 23 Brown, A. M., Hoopes, S. L., White, R. H. & Sarisky, C. A. Purine biosynthesis in archaea: variations on a theme. *Biology direct* **6**, 63, doi:10.1186/1745-6150-6-63 (2011).
- 24 Oro, J., Yamrom, T. & Cortes, S. The Prebiotic Synthesis of Imidazoles and Their Catalytic Role in the Synthesis of Oligonucleotides. *Fed Proc* **42**, 1926-1926 (1983).
- 25 Basile, B., Lazcano, A. & Oro, J. Prebiotic syntheses of purines and pyrimidines. *Adv Space Res* **4**, 125-131 (1984).
- 26 Meléndez-Hevia, E., Waddell, T. G. & Montero, F. Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle. *J Theor Biol* **166**, 201-220 (1994).
- 27 Morowitz, H., Peterson, E. & Chang, S. The Synthesis of Glutamic-Acid in the Absence of Enzymes - Implications for Biogenesis. *Origins Life Evol B* **25**, 395-399, doi:Doi 10.1007/Bf01581777 (1995).
- 28 Waddell, T. G., Eilders, L. L., Patel, B. P. & Sims, M. Prebiotic methylation and the evolution of methyl transfer reactions in living cells. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **30**, 539-548 (2000).
- 29 Wixon, J. & Kell, D. The Kyoto encyclopedia of genes and genomes--KEGG. *Yeast* **17**, 48-55, doi:10.1002/(SICI)1097-0061(200004)17:1<48::AID-YEA2>3.0.CO;2-H (2000).
- 30 Caspi, R. *et al.* MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic acids research* **34**, D511-516, doi:10.1093/nar/gkj128 (2006).
- 31 Claudel-Renard, C., Chevalet, C., Faraut, T. & Kahn, D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic acids research* **31**, 6633-6639 (2003).
- 32 Bairoch, A. The ENZYME database in 2000. *Nucleic acids research* **28**, 304-305 (2000).
- 33 Mavromatis, K. *et al.* The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Standards in Genomic Sciences* **1**, 63 (2009).
- 34 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-3402 (1997).
- 35 Wilson, D., Madera, M., Vogel, C., Chothia, C. & Gough, J. The SUPERFAMILY database in 2007: families and functions. *Nucleic acids research* **35**, D308-313, doi:10.1093/nar/gkl910 (2007).
- 36 Lo Conte, L. *et al.* SCOP: a structural classification of proteins database. *Nucleic acids research* **28**, 257-259 (2000).
- 37 Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**, 127-136, doi:10.1038/nrmicro751 (2003).
- 38 Vantrump, J. E. & Miller, S. L. Carbon-Monoxide on Primitive Earth. *Earth Planet Sc Lett* **20**, 145-150, doi:Doi 10.1016/0012-821x(73)90152-0 (1973).

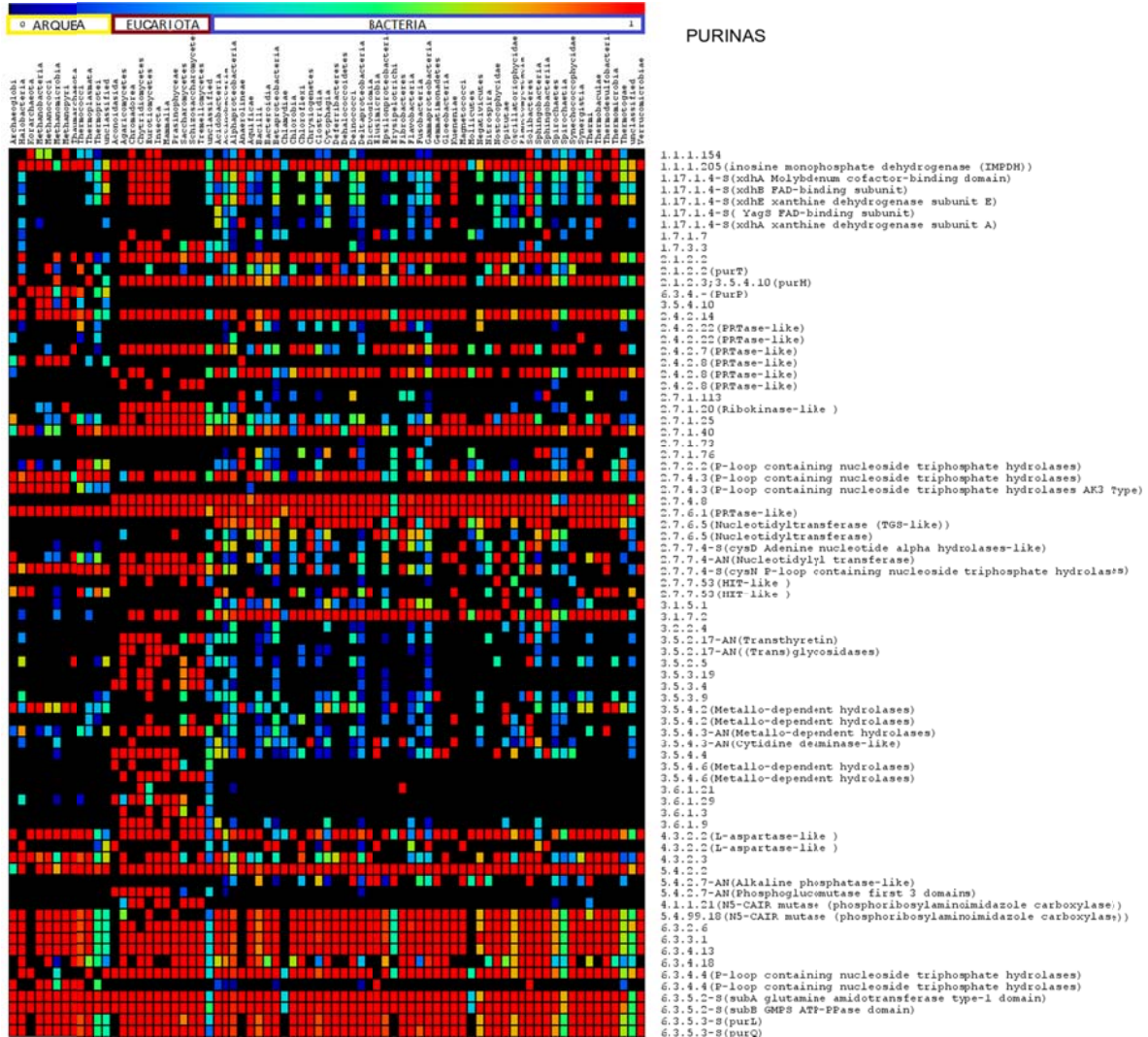
- 39 Smirnov, A. & Schoonen, M. A. Evaluating experimental artifacts in hydrothermal prebiotic synthesis experiments. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life* **33**, 117-127 (2003).
- 40 Martin, W. & Russell, M. J. On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **362**, 1887-1925, doi:10.1098/rstb.2006.1881 (2007).
- 41 Wallace, K. B. & Starkov, A. A. Mitochondrial targets of drug toxicity. *Annual review of pharmacology and toxicology* **40**, 353-388, doi:10.1146/annurev.pharmtox.40.1.353 (2000).
- 42 Nicholls, P. Formate as an Inhibitor of Cytochrome-C Oxidase. *Biochem Bioph Res Co* **67**, 610-616, doi:Doi 10.1016/0006-291x(75)90856-6 (1975).
- 43 Haritos, V. S. & Dojchinov, G. Cytochrome c oxidase inhibition in the rice weevil *Sitophilus oryzae* (L.) by formate, the toxic metabolite of volatile alkyl formates. *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP* **136**, 135-143 (2003).
- 44 Tribunskikh, I. A., Alenin, V. V., Selivanov, S. I., Shavva, A. G. & Inge-Vechtomov, S. G. Divergence of de novo biosynthesis of inosine-5'-triphosphate. *Doklady. Biochemistry and biophysics* **400**, 65-68 (2005).
- 45 Dismukes, G. *et al.* The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences* **98**, 2170-2175 (2001).
- 46 Grotzinger, J. P. & Kasting, J. F. New Constraints on Precambrian Ocean Composition. *Journal of Geology* **101**, 235-243 (1993).
- 47 Meyer, E., Kappock, T. J., Osuji, C. & Stubbe, J. Evidence for the direct transfer of the carboxylate of N5-carboxyaminoimidazole ribonucleotide (N5-CAIR) to generate 4-carboxy-5-aminoimidazole ribonucleotide catalyzed by *Escherichia coli* PurE, an N5-CAIR mutase. *Biochemistry-Us* **38**, 3012-3018, doi:10.1021/bi9827159 (1999).
- 48 Cusack, N. J., Shaw, G. & Litchfield, G. J. Purines, pyrimidines, and imidazoles. XXXVI. Carboxylation of some 5-aminoimidazoles and related compounds, including nucleosides and nucleotides, with potassium hydrogen carbonate in aqueous solution. *Journal of the Chemical Society. Perkin transactions 1* **8**, 1501-1507 (1971).
- 49 Cordat, E. & Casey, J. R. Bicarbonate transport in cell physiology and disease. *Biochemical Journal* **417**, 423-439, doi:Doi 10.1042/Bj20081634 (2009).
- 50 Grochowski, L. L., Censky, K., Xu, H. M. & White, R. H. A new class of adenylate kinase in methanogens is related to uridylate kinase. *Arch Microbiol* **194**, 141-145, doi:DOI 10.1007/s00203-011-0759-9 (2012).
- 51 Yan, H. & Tsai, M. D. Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. *Advances in enzymology and related areas of molecular biology* **73**, 103-134, x (1999).
- 52 Stolworthy, T. S. & Black, M. E. The mouse guanylate kinase double mutant E72Q/D103N is a functional adenylate kinase. *Protein engineering* **14**, 903-909 (2001).
- 53 Feng, X., Yang, R. N., Zheng, X. F. & Zhang, F. Y. Identification of a novel nuclear-localized adenylate kinase 6 from *Arabidopsis thaliana* as an essential stem growth factor. *Plant Physiol Bioch* **61**, 180-186, doi:DOI 10.1016/j.plaphy.2012.10.002 (2012).
- 54 Ren, H. *et al.* The crystal structure of human adenylate kinase 6: An adenylate kinase localized to the cell nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 303-308, doi:10.1073/pnas.0407459102 (2005).
- 55 Keegan, L. P., Leroy, A., Sproul, D. & O'Connell, M. A. Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. *Genome biology* **5**, 209, doi:10.1186/gb-2004-5-2-209 (2004).

- 56 Hitchings, G. H. The purine metabolism of protozoa. *Advances in Enzyme Regulation* **20**, 375-386 (1982).
- 57 Zubay, G. To what extent do biochemical pathways mimic prebiotic pathways. *Chemtracts Biochem Mol Biol* **4**, 317-323 (1993).
- 58 Frank, R. A., Leeper, F. J. & Luisi, B. F. Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cellular and molecular life sciences : CMLS* **64**, 892-905, doi:10.1007/s00018-007-6423-5 (2007).
- 59 Groziak, M. P., Bhat, B. & Leonard, N. J. Nonenzymatic synthesis of 5-aminoimidazole ribonucleoside and recognition of its facile rearrangement. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7174-7176 (1988).
- 60 Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952-956, doi:10.1038/nature01145 (2002).
- 61 Sakumi, K. *et al.* ITPA protein, an enzyme that eliminates deaminated purine nucleoside triphosphates in cells. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **703**, 43-50 (2010).
- 62 Glasemacher, J., Bock, A. K., Schmid, R. & Schönheit, P. Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. *European journal of biochemistry / FEBS* **244**, 561-567 (1997).
- 63 Krepl, M., Otyepka, M., Banas, P. & Sponer, J. Effect of Guanine to Inosine Substitution on Stability of Canonical DNA and RNA Duplexes: Molecular Dynamics Thermodynamics Integration Study. *J Phys Chem B* **117**, 1872-1879, doi:Doi 10.1021/Jp311180u (2013).
- 64 Rowland, P., Norager, S., Jensen, K. F. & Larsen, S. Structure of dihydroorotate dehydrogenase B: electron transfer between two flavin groups bridged by an iron-sulphur cluster. *Structure* **8**, 1227-1238 (2000).
- 65 Takashima, E. *et al.* Characterization of the dihydroorotate dehydrogenase as a soluble fumarate reductase in *Trypanosoma cruzi*. *Mol Biochem Parasit* **122**, 189-200, doi:Pii S0166-6851(02)00100-7
- Doi 10.1016/S0166-6851(02)00100-7 (2002).
- 66 Norager, S., Jensen, K. F., Bjornberg, O. & Larsen, S. E-coli dihydroorotate dehydrogenase reveals structural and functional distinctions between different classes of dihydroorotate dehydrogenases. *Structure* **10**, 1211-1223, doi:Pii S0969-2126(02)00831-6
- Doi 10.1016/S0969-2126(02)00831-6 (2002).
- 67 Cavalier-Smith, T. Cell evolution and Earth history: stasis and revolution. *Philos T R Soc B* **361**, 969-1006, doi:DOI 10.1098/rstb.2006.1842 (2006).
- 68 Valas, R. E. & Bourne, P. E. The origin of a derived superkingdom: how a gram-positive bacterium crossed the desert to become an archaeon. *Biology direct* **6**, doi:Artn 16
- Doi 10.1186/1745-6150-6-16 (2011).
- 69 Blobel, G. Intracellular protein topogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 1496-1500 (1980).
- 70 Myllykallio H, S. S., Grosjean H. in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (ed Grosjean Henri) 275-288 (Landes Bioscience, Austin (TX), 2009).
- 71 Escartin, F., Skouloubris, S., Liebl, U. & Myllykallio, H. Flavin-dependent thymidylate synthase X limits chromosomal DNA replication. *Proceedings of the National Academy of*

- Sciences of the United States of America* **105**, 9948-9952, doi:10.1073/pnas.0801356105 (2008).
- 72 Myllykallio, H. *et al.* An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* **297**, 105-107, doi:10.1126/science.1072113 (2002).
- 73 Chan, C. X., Bhattacharya, D. & Reyes-Prieto, A. Endosymbiotic and horizontal gene transfer in microbial eukaryotes: Impacts on cell evolution and the tree of life. *Mobile genetic elements* **2**, 101-105, doi:10.4161/mge.20110 (2012).
- 74 Nygaard, P. & Munch-Petersen, A. (Academic Press, London, 1983).
- 75 Bick, M. D. & Davidson, R. L. Total substitution of bromodeoxyuridine for thymidine in the DNA of a bromodeoxyuridine-dependent cell line. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 2082-2086 (1974).
- 76 Warren, R. A. Modified bases in bacteriophage DNAs. *Annual review of microbiology* **34**, 137-158, doi:10.1146/annurev.mi.34.100180.001033 (1980).
- 77 Marliere, P. *et al.* Chemical evolution of a bacterium's genome. *Angew Chem Int Ed Engl* **50**, 7109-7114, doi:10.1002/anie.201100535 (2011).

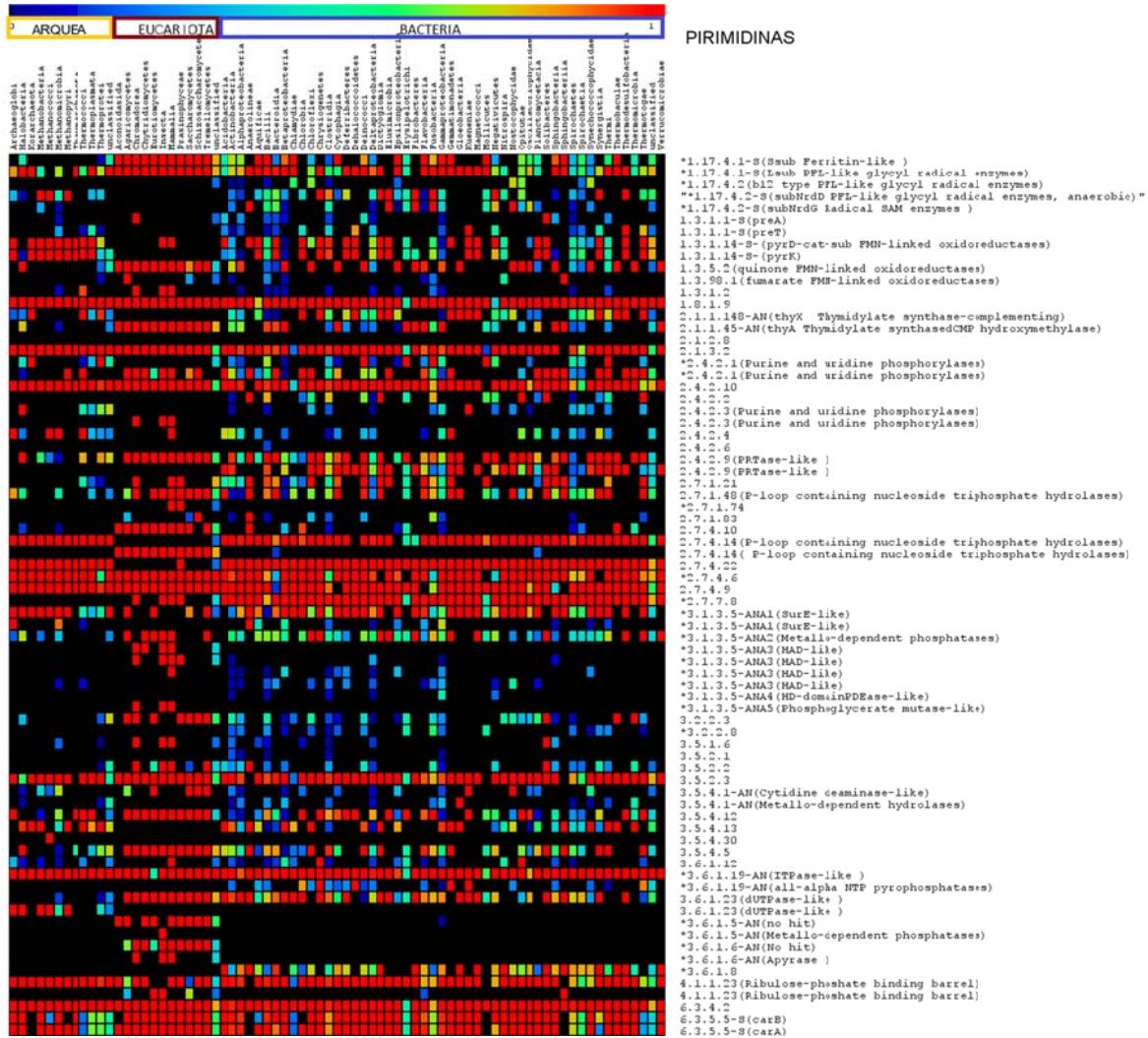
13-APÉNDICE

13.1- Distribución taxonómica en la biosíntesis de purinas (S1)



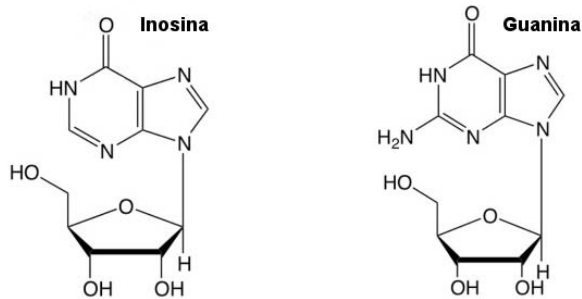
Promedio de la distribución taxonómica de las enzimas del metabolismo de purinas en los tres dominios celulares. La distribución taxonómica de las familias enzimáticas (etiquetas verticales) fue calculada buscando su distribución de ortólogos a través de los diversos grupos taxonómicos (Arque, Bacteria y Eucaria, etiquetas horizontales). Las enzimas se encuentran ordenadas de acuerdo a su número enzimático. Algunos números enzimáticos redundantes con asignación de superfamilia idéntica son complementarios de acuerdo con la base de datos PRIAM. La notación "S" denota subunidad proteica; "AN" enzimas análogas, aquellas con el mismo número enzimático pero que pertenecen a diferentes superfamilias estructurales, sugiriendo diferente origen evolutivo. Los colores en escala de arcoiris, donde todas aquellas enzimas en azul oscuro mantienen valores promedio de (0), mientras que el rojo denota altos valores (1).

13.2- Distribución taxonómica en la biosíntesis de pirimidinas (S2)



Distribución taxonómica promedio de las enzimas pertenecientes al metabolismo de pirimidinas. Las etiquetas y colores representan los mismos que en S1. En asterisco son indicadas todas aquellas enzimas comunes al metabolismo de purinas.

13.3- Similitud estructural entre la Inosina y Guanina (S3)



A la izquierda de la figura se puede observar la inosina con gran similitud estructural a la guanina (a la derecha de la figura).

13.4- Glosario

AICA: El 5-amino-4-imidazol carboxiamida (AICA), molécula precursora de nucleótidos, la cual se puede sintetizar en condiciones consideradas prebióticas.

DOMINIO PROTEICO: Es un término genérico que designa una región de una proteína con interés biológico funcional o estructural. También se llama dominio a una región de la estructura tridimensional de una proteína con una función concreta, que incluye regiones no necesariamente contiguas en la secuencia de aminoácidos.

DOMINIOS CELULARES: El sistema de tres dominios es una clasificación biológica propuesta por Carl Woese y colaboradores en 1977, que clasifica el árbol de la vida en tres grupos primarios: Bacteria, Arquea y Eucaria, Woese se basó en las diferencias encontradas en la secuencia del ARN ribosomal de la subunidad menor.

DUPLICACIÓN GÉNICA: Es la repetición de un fragmento de cromosoma. Las duplicaciones surgen por error en la duplicación del ADN, como producto de una reorganización cromosómica de tipo estructural o relacionado con un proceso de sobrecruzamiento defectuoso. Las duplicaciones no suelen ser deletéreas, son una fuente de nuevo material genético y base para nuevos cambios evolutivos. Muchas de las familias génicas con un origen evolutivo común, o las familias multigénicas pueden tener su origen en las duplicaciones.

FAMILIA ESTRUCTURAL: Los dominios en una superfamilia son agrupados en familias, los cuales mantienen un ancestro en común más reciente.

FOLATOS: Moléculas que conforman las vitaminas como la B9. La B9 ayuda a convertir la vitamina B12 en una de sus formas coenzimáticas y participa en la síntesis de ADN requerido para un rápido crecimiento celular. Del mismo modo actúa como coenzima en la transferencia de grupos monocarbonados.

FOLD (PLEGAMIENTO): Las diferentes formas tridimensionales de un dominio estructural.

FUMARATO: Sal del ácido fumárico, un intermediario clave en el ciclo de Krebs.

KEGG: (Enciclopedia de Genes y Genomas de Kyoto) es una colección de bases de datos en línea de genomas, rutas enzimáticas, y químicos biológicos.

LCA: El último ancestro común de todos los organismos vivos actuales.

NAD: nicotinamida adenina dinucleótido, está formada por dos nucleótidos unidos a través de sus grupos fosfatos, siendo uno de ellos una base de adenina y el otro de nicotinamida. Su función principal es el intercambio de electrones e hidrogeniones en la producción de energía de todas las células.

NADH: Forma reducida del nicotinamida adenina dinucleótido.

NUCLEOSIDO: El nucleósido es la parte del nucleótido formada únicamente por la base nitrogenada y la pentosa.

NUCLEOTIDO: Los nucleótidos son moléculas orgánicas formadas por la unión covalente de un monosacárido de cinco carbonos (pentosa), una base nitrogenada y un grupo fosfato.

ORGANISMOS GRAM –: Aquellas bacterias que NO se tiñen de azul oscuro o violeta por la tinción de Gram. La envoltura celular de las bacterias Gram-negativas está compuesta por una membrana citoplasmática (membrana interna), una pared celular delgada de peptidoglicano, que rodea a la anterior, y una membrana externa que recubre la pared celular de estas bacterias. De cualquier manera, aunque la tinción en gram es un criterio empírico, sus bases descansan en una marcada diferencia en la composición química y ultra estructura de la pared celular bacteriana, delimitada por la presencia y ausencia de la membrana lipídica externa. Este criterio no es infalible ya que un número de bacterias que poseen un sola membrana pueden teñir como gram negativas debido a la pérdida de la capa de peptidoglicano o por su composición de la pared celular aunque muestren una relación cercana con gram-positivas y viceversa algunas bacterias que son positivas en la tinción de

gram debido a una gruesa capa de peptidoglicano también pueden poseer una membrana celular externa.

ORGANISMOS GRAM +: aquellas bacterias que se tiñen de azul oscuro o violeta por la tinción de Gram. La capa de peptidoglicano confiere una gran resistencia a estas bacterias y es la responsable de retener el tinte durante la tinción de Gram. Realmente, no todas las bacterias del grupo son Gram-positivas (no se tiñen por la aplicación de ese método), pero se incluyen aquí por su similitud molecular con otras bacterias Gram-positivas

PIRIDOXAL FOSFATO: (PLP, piridoxal-5'-fosfato, P5P) es un grupo prostético ligado a ciertas enzimas. El PLP es también la forma activa de la vitamina B₆. El PLP actúa como coenzima en todas las reacciones de transaminación, y en ciertas reacciones de descarboxilación, desaminación, y racemización de aminoácidos.

QUÍMICA PREBIÓTICA: Encaminada a conseguir unas condiciones experimentales que reprodujesen aquéllas en las que pudo surgir la vida, y averiguar los compuestos químicos que podrían sintetizarse en dichas condiciones y si era factible así el origen de la vida.

MODELO OCULTO DE MÁRKOV: En un modelo de Márkov normal, el estado es visible directamente para el observador, por lo que las probabilidades de transición entre estados son los únicos parámetros. En un modelo *oculto* de Márkov, el estado no es visible directamente, sino que sólo lo son las variables influidas por el estado. Cada estado tiene una distribución de probabilidad sobre los posibles símbolos de salida. Consecuentemente, la secuencia de símbolos generada por un HMM proporciona cierta información acerca de la secuencia de estados.

RPS-BLAST (Reversed Position Specific Blast): Busca secuencias contra una base de datos de perfiles (pssm). Esto es opuesto al PSI-BLAST que busca perfiles en una base de datos de secuencias.

RPS-BLAST PERFILES: Son pssm, o matrices de puntuación posición específica, es un tipo de matriz de puntuación usada en BLAST, en el cual la puntuación en la sustitución de aminoácidos es dada por separado para cada posición en un alineamiento múltiple de secuencias de proteínas.

RIBOTSWITCHES: Elementos de ARNm que unen metabolitos o iones como ligandos, para regular procesos de expresión mediante la formación de estructuras alternativas.

RIBOZIMAS: Las ribozimas son ARN con actividad catalítica. El término "ribozima" es una contracción de las palabras "ácido ribonucleico" y "enzima". Los sustratos de las ribozimas son con frecuencia ARN. La actividad de las ribozimas combina transesterificación e hidrólisis de un enlace fosfodiéster.

SCOP: Base de datos de la clasificación estructural de proteínas (SCOP) consta de una amplia clasificación manual de los dominios estructurales de proteínas basadas en similitudes estructurales y de secuencia. La motivación para esta clasificación es determinar las relaciones evolutivas entre proteínas. Proteínas con la misma forma pero poca similitud de secuencia o funcional son ubicadas en diferentes superfamilias y se asume que pudieron haber tenido un ancestro lejano en común.

STARTER TYPES ENZYMES: Enzimas iniciales en español. Es referida en la teoría semi enzimática del origen de las rutas metabólicas por Lazcano-Miller. Son todas aquellas enzimas que surgieron a partir de reacciones no enzimáticas, utilizando como templado los mismos intermediarios de las reacciones no enzimáticas.

SUPERFAMILIA ESTRUCTURAL: Los dominios en un plegamiento son agrupados en una superfamilia los cuales mantienen al menos un ancestro común distante.

SUPERFAMILY: Es una base de datos que clasifica secuencias de aminoácidos en dominios estructurales, específicamente en superfamilias clasificadas por la base de datos SCOP. Las superfamilias son grupos de proteínas las cuales tiene evidencia estructural para soportar un ancestro evolutivo en común, aunque tal vez no poseen homología detectable a nivel secuencia.

SUSTRATO: es una molécula sobre la que actúa una enzima.

TEORÍA DEL MUNDO RNA: La hipótesis del mundo del ARN es apoyada por la observación de que muchos de los componentes más críticos de las células están compuestos mayormente o completamente de ARN. Esto significaría que el ARN en las células modernas es un vestigio de la evolución del mundo de ARN que nos precedió. También, muchos cofactores críticos son o bien nucleótidos o sustancias claramente relacionados con ellos.

TEORÍA ENDOSIMBIOTICA: explicar el proceso por el cual las células procariotas evolucionaron hacia eucariotas, adquiriendo sus orgánulos (como son los cloroplastos y las mitocondrias), proceso conocido como la teoría del origen endosimbiótico de los eucariotas.

TIAMINA: También conocida como vitamina B1. Es una molécula que consta de 2 estructuras cíclicas orgánicas interconectadas: un anillo de pirimidina con un grupo amino y un anillo tiazol azufrado unido a la pirimidina por un puente metileno. Actúa como coenzima en el metabolismo.

14-ANEXO

14.1 Artículo publicado en BMC Genomics

RESEARCH ARTICLE

Open Access

Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life

Dagoberto Armenta-Medina^{1*}, Lorenzo Segovia¹ and Ernesto Perez-Rueda^{1,2*}

Abstract

Background: Nucleotide metabolism is central to all biological systems, due to their essential role in genetic information and energy transfer, which in turn suggests its possible presence in the last common ancestor (LCA) of *Bacteria*, *Archaea* and *Eukarya*. In this context, elucidation of the contribution of the origin and diversification of *de novo* and salvage pathways of nucleotide metabolism will allow us to understand the links between the enzymatic steps associated with the LCA and the emergence of the first metabolic pathways.

Results: In this work, the taxonomical distribution of the enzymes associated with nucleotide metabolism was evaluated in 1,606 complete genomes. 151 sequence profiles associated with 120 enzymatic reactions were used. The evaluation was based on profile comparisons, using RPS-Blast. Organisms were clustered based on their taxonomical classifications, in order to obtain a normalized measure of the taxonomical distribution of enzymes according to the average of presence/absence of enzymes per genus, which in turn was used for the second step, to calculate the average presence/absence of enzymes per Clade.

Conclusion: From these analyses, it was suggested that divergence at the enzymatic level correlates with environmental changes and related modifications of the cell wall and membranes that took place during cell evolution. Specifically, the divergence of the 5-(carboxyamino) imidazole ribonucleotide mutase to phosphoribosylaminoimidazole carboxylase could be related to the emergence of multicellularity in eukaryotic cells. In addition, segments of salvage and *de novo* pathways were probably complementary in the LCA to the synthesis of purines and pyrimidines. We also suggest that a large portion of the pathway to inosine 5'-monophosphate (IMP) in purines could have been involved in thiamine synthesis or its derivatives in early stages of cellular evolution, correlating with the fact that these molecules may have played an active role in the protein-RNA world. The analysis presented here provides general observations concerning the adaptation of the enzymatic steps in the early stages of the emergence of life and the LCA.

Keywords: Nucleotide metabolism, Comparative genomics, Evolution, Sequence profiles, Enzymes, LCA

Background

Metabolism represents an intricate ensemble of enzyme-catalyzed reactions that lead to synthesis and degradation of compounds within the cell. In recent years, an increasing amount of information on metabolism from different species has become available, allowing for comparative genomic-scale studies on the evolution of specific pathways or whole metabolic networks [1-4]. Metabolism can be considered one of the most ancient biological

networks, where nodes represent substrates or enzymes and edges represent the relationships between them. From this perspective, the study of metabolic networks is focused on describing topological properties, such as the existence of functional modules, giving special relevance to clustering and motif formation, and showing the existence of similar attributes to the small-world and scale-free networks [3].

Therefore, two main hypotheses on the origin and evolution of enzyme-driven metabolism have been postulated

* Correspondence: armentamedina@gmail.com; erueda@ibt.unam.mx

¹Departamento de Ingeniería Celular y Biotecnología, Instituto de Biotecnología, UNAM Av. Universidad 2001, Cuernavaca, Morelos CP 62210, México
Full list of author information is available at the end of the article

based on the notion that gene duplication, followed by divergence, can lead to the origin of new metabolic reactions. The “stepwise hypothesis” [5] suggests that, in the case where a substrate tends to be depleted, gene duplication can provide an enzyme capable of supplying the exhausted substrate, giving rise to homologous enzymes that catalyze consecutive reactions. On the other hand, the “patchwork hypothesis” [6] proposes that duplication of genes encoding promiscuous enzymes (capable of catalyzing multiple reactions) allows each descendant enzyme to specialize in one of the ancestral reactions.

In this regard, it is plausible that a small number of enzymes with broad specificity existed in early stages of metabolic evolution. Genes encoding these enzymes would have been duplicated, generating enzymes that, through sequence divergence, became more specialized [7].

Collectively, these studies have highlighted the contribution of gene duplication in the evolution of metabolism [4]. In recent works, the universal occurrence of some pathways and branches, such as diverse amino acid pathways, in modern species suggests that they existed in the last common ancestor (LCA) of *Bacteria*, *Archaea* and *Eukarya* [1,8]. However, despite the importance of nucleotide metabolism in all organisms, few studies have addressed this issue by using genomic approaches [9,10]. Nucleotide metabolism is central in all living systems, due to its role in transferring genetic information and energy. Indeed, it has been described as one of the ancient metabolisms in evolution. Specifically, the emergence of an ancestral folding or P-loop hydrolase appeared parallel to this metabolism [11], reinforcing its antiquity. In addition, many of the intermediates associated with this metabolic module have been intimately associated with prebiotic chemistry and the origin of life [10,12,13]. In this regard, we adopted a multigenomic strategy for the reconstruction and analysis of the metabolism of nucleotides, evaluating the contribution of the origin and diversification of *de novo* and salvage pathways for nucleotides in the evolution of organisms. In addition, these analyses allow the identification of a metabolic link between the LCA and the first steps in the structure of biological networks [14-16]. Our strategy reveals some general rules concerning the adaptation of the first predominant chemical reactions to enzymatic steps in the LCA and allows us to infer environmental issues in the early stages of the emergence of life. In addition, it was possible to determine the presence of *de novo* biosynthesis pathways of the ribonucleotides and deoxyribonucleotides of uracil and cytosine for pyrimidine metabolism associated with the LCA. Finally, we found differences in the enzymes for nucleotide metabolism that correlated with environmental changes and with associated cellular architecture adaptations; such is the case for the enzymes involved in the synthesis of 5-aminoimidazole ribonucleotide (AIR) to

(4-carboxyaminoimidazole ribonucleotide) (CAIR). These findings suggest the presence of HCO_3^- in primitive seas and its use as one of the main carbon sources for the first organisms. Our main results derived from the taxonomical distribution of enzymatic families belonging to nucleotide metabolism, supported by experimental evidence, are described in this report.

Results and discussion

Taxonomic distribution of nucleotide metabolic enzymes

The taxonomic distribution of proteins provides clues concerning the relative occurrence of the enzymes and their branches and paths for the evolution of metabolism [1,8,11,17,18]. In this regard, the origin and evolution of nucleotide metabolism was traced in organisms belonging to the three cellular domains, *Bacteria*, *Archaea* and *Eukarya*, by an exhaustive evaluation of the taxonomic distribution of their enzymatic repertoires. Therefore, each enzymatic activity encoded by an EC number can exhibit different profiles, which are individual vectors of the presence/absence of homologous enzymes in all the genomes. In total, 151 profiles associated with 120 enzymatic reactions related to nucleotide metabolism were used to scan 1,606 genomes by using RPS-BLAST. (Additional file 1: Figure S1 and Additional file 2: Figure S2 and Additional file 3: Tables S1 and Additional file 4: Table S2). Based on these comparisons, evolutionary origins of nucleotide metabolism can be traced, even close to the LCA of all organisms. In this regard, enzymes widely distributed across the three cellular domains are proposed to have been present in the LCA [8,11,18-20]. Alternatively, enzymes constrained to specific clades or cellular domains would suggest adaptations of organisms or cellular domains to specific lifestyles. Based on these considerations, we discuss our most notable results in the following sections.

Evolution of purine metabolism

De novo purine biosynthesis

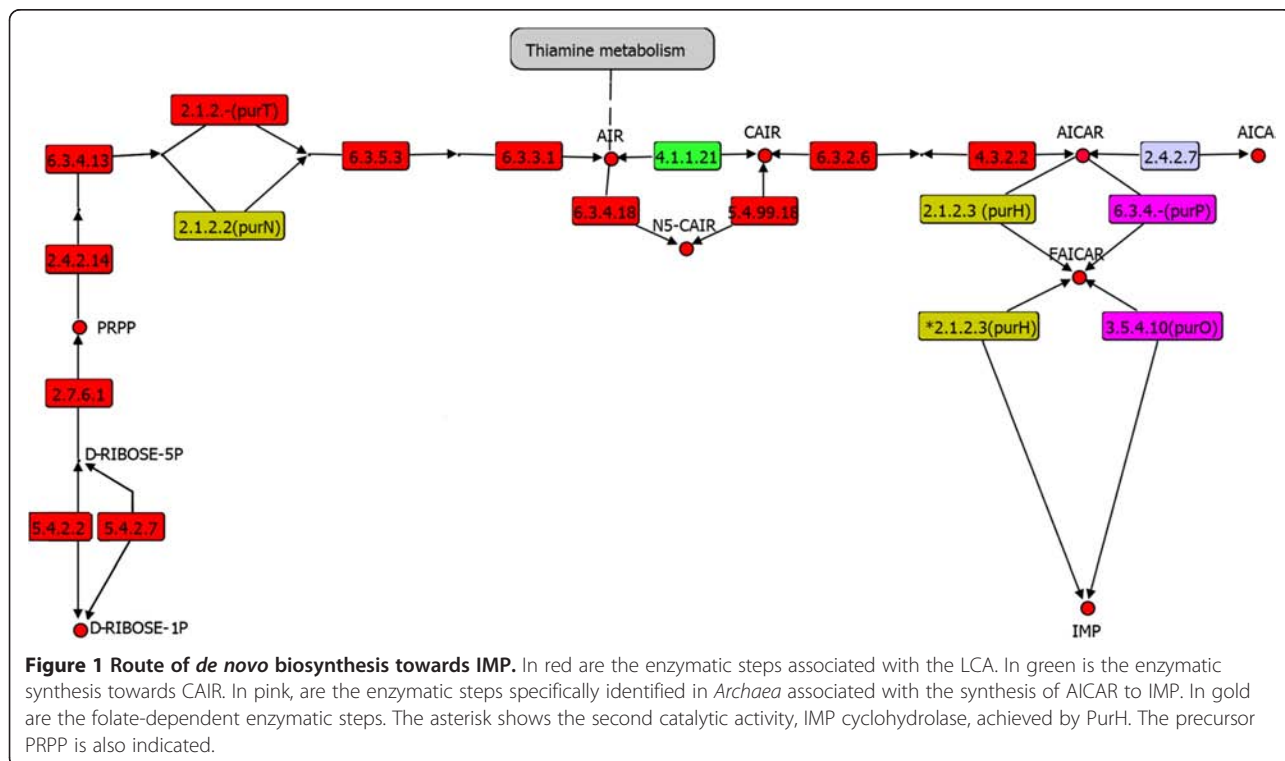
The *de novo* biosynthesis of purines, starting from D-ribose-1-phosphate to inosine 5'-monophosphate (IMP) production, the main intermediate in the synthesis of ribonucleotides and deoxyribonucleotides, guanine and adenine, follows a linear branch. The first step is associated with phosphoglucomutase (EC 5.4.2.2) or phosphopentomutase (5.4.2.7) and the second step is associated with ribose-phosphate diphosphokinase (2.7.6.1); both steps are necessary for the synthesis of 5-phospho-alpha-D-ribose-1-pyrophosphate (PRPP), which starts from D-ribose-1-phosphate. Based on their taxonomical distribution, the enzymes associated with the 5.4.2.2 and 2.7.6.1 catalytic activities were identified as being widely distributed among *Bacteria*, *Archaea* and *Eukarya*, suggesting the

probable existence of PRPP biosynthesis in the LCA. Indeed, PRPP is a key precursor for biosynthesis in the *de novo* and salvage pathways for purines and pyrimidines; however, this intermediary is unstable and susceptible to hydrolysis. Therefore, it is probable that its abiotic synthesis, if it occurred, was not enough to maintain the biosynthesis in the LCA [10].

Therefore, the first step for purine biosynthesis, the catalysis to ribose-5-phosphate starting from ribose 1-phosphate, is achieved by either of the two enzymes related to the enzymes EC 5.4.2.2 and 5.4.2.7. These two enzymes are analogous, since no homology at the sequence or structural level was detected. The enzyme 5.4.2.7 is partially distributed in *Bacteria*, mainly in free-living organisms associated with a host, such as *Streptococcus pneumoniae* and *Lactobacillus rhamnosus*; however, it was not found in archaeal and eukaryal organisms, suggesting that its emergence was posterior to the LCA divergence, probably as a secondary adaptation associated with the bacterial host.

Starting from the intermediary PRPP, in the linear branch towards IMP biosynthesis, we identified enzymes belonging to five catalytic steps (amidophosphoribosyltransferase, 2.4.2.14; phosphoribosylamine-glycine ligase, 6.3.4.13; phosphoribosylglycinamide formyltransferase, 2.1.2.2; phosphoribosylformylglycinamide synthase, 6.3.5.3; phosphoribosylformylglycinamide cyclo-ligase, 6.3.3.1) required for the transformation of PRPP into AIR [21]. Most of these enzymes were identified as widely distributed in the three cellular domains, suggesting

their presence in the LCA (Figure 1). The enzymatic step associated with EC 2.1.2.2 is responsible for the transformation of glycinamide ribotide (GAR) to formylglycinamide ribotide (FGAR) and could be carried out by two enzymes associated with different evolutionary families, PurN (Figure 1 Gold box) or phosphoribosylglycinamide formyltransferase and PurT or phosphoribosylglycinamide formyltransferase 2 [9]. Proteins associated with the PurN family use derivatives from folate synthesis as substrates. This family was identified as widely distributed in *Bacteria* and *Eukarya* and partially distributed in *Archaea*. Alternatively, proteins from the PurT family were partially distributed in *Archaea* and *Bacteria* and sparsely in *Eukarya*. It is probable that the PurT enzymatic family could have been present in the LCA, with posterior loss events in *Eukarya* due to its requirement for formate as a substrate. In this regard, formate is described as one-carbon donor and one of the main molecules present in prebiotics conditions, prior to folate metabolism [22-24]. In a posterior phase, the emergence of folate biosynthesis might have facilitated the emergence of PurN (Figure 1, gold box), thereby achieving the co-occurrence of PurN and PurT in the LCA. Indeed, previous works have suggested that the emergence of PurT preceded the emergence of PurN, mainly because PurT utilizes a more primitive substrate prior to the folate-dependent pathway [24]. One of the evolutionary pressures for the selection of PurN instead of PurT in eukaryotic organisms could be associated with the emergence of the mitochondrial respiratory chain. It



has been shown that the PurT substrate, formate, is toxic and binds to cytochrome *c* oxidase-like [25,26], uncoupling the redox reactions and favoring the selection of PurN in eukaryotic organisms.

Another enzymatic step identified in the purine biosynthetic pathway is achieved by the phosphoribosylformylglycinamide synthase (6.3.5.3) that transforms FGAR to formylglycinamide-ribonucleotide (FGAM). This enzyme is composed of two catalytic subunits, PurQ and PurL. In general, this enzyme may occur as a multidomain protein or as a multiprotein complex, where the subunits L and Q are encoded independently. Altogether, we identified a co-occurrence of these subunits in the three cellular domains. Finally, the last step towards AIR transformation is carried out by the cyclo-ligase phosphoribosylformylglycinamide (EC 6.3.3.1), which transforms FGAM to AIR, and in turn this enzyme was found to be widely distributed in the three cellular domains, suggesting its presence in the LCA.

The product of the steps previously described, the AIR 5-amino-1-(5-phospho-D-ribosyl) imidazole-4-carboxylate, can be transformed by two pathways; in the first one, only one reaction, catalyzed by 4.1.1.21 (AIR carboxylase) (Figure 1, green box), is involved in CAIR synthesis; in the second, a two-step pathway, entails two reactions involving N5-carboxyaminoimidazole ribonucleotide synthase (6.3.4.18) and 5-(carboxyamino) imidazole ribonucleotide mutase (5.4.99.18). When the taxonomical distribution of these enzymes was evaluated, it was found that the N5-carboxyaminoimidazole ribonucleotide synthase (6.3.4.18) was widely distributed among all the organisms in the three cellular domains, whereas the enzyme 5-(carboxyamino) imidazole ribonucleotide mutase (5.4.99.18) exhibited a wide distribution in *Bacteria* and *Archaea* and in some eukaryal clades, suggesting their ancestry in all the organisms with posterior gene loss in *Eukarya* (Figure 1). Additionally, these enzymes were analyzed at the sequence level, where the 5-(carboxyamino) imidazole ribonucleotide mutase (5.4.99.18) was evolutionarily related to AIR carboxylase (4.1.1.21) [27]. Indeed, at the structural level both enzymes share a common evolutionary origin, belonging to the N5-CAIR mutase (phosphoribosylaminoimidazole carboxylase superfamily). The observation that AIR carboxylase (4.1.1.21) is almost exclusively distributed in the *Eukarya* suggests its posterior subfunctionalization from 5-(carboxyamino) imidazole ribonucleotide mutase (5.4.99.18) [28].

The discovery of the emergence of these branches is interesting, because it allows us to infer the processes associated with environmental changes on Earth. In this context, Tribunskikh et al. [29] suggested that the divergence of the phosphoribosylaminoimidazole carboxylase (4.1.1.21) to 5-(carboxyamino) imidazole ribonucleotide mutase (5.4.99.18) was a consequence of decreasing atmospheric

CO₂, which resulted in the addition of the 5-(carboxyamino) imidazole ribonucleotide synthase (6.3.4.18) and a change of specificity and, by consequence, in the two-step pathway. Those authors supported their proposal by the fact that the two-step pathways could work at low CO₂ levels, under aerobic or anaerobic conditions. However, our data suggest an alternative scenario where, although the concentration of CO₂ in the primitive atmosphere was high, acidic conditions in early oceans favored the formation of HCO₃⁻. This is consistent with simulations showing that the concentration of HCO₃⁻ oscillated between levels 30 to 30,000 times higher in the early oceans than current levels [30]. Interestingly, the enzyme 6.3.4.18, which is associated with the two-step pathway, uses HCO₃⁻ as a substrate, and HCO₃⁻ is considered a dominant form of CO₂ in early oceans [30,31], which together with our data of taxonomic distribution supports the notion that this pathway preceded the path where the enzymatic reaction catalyzed by 4.1.1.21 uses CO₂ as the substrate. Indeed, it has been experimentally shown that in aqueous solutions with high concentrations of KHCO₃, AIR is easily converted into CAIR in the absence of enzymes [29,32,33]. This mechanism could take place through the accumulation of the intermediate N-CAIR (N5-carboxy-AIR, or N5-carboxyaminoimidazole ribonucleotide), which then undergoes a rearrangement to CAIR. These reactions appear to be the template whereby the enzymatic activities adapted in the two-step pathway, widespread in the cellular domains and probably in the LCA. Probably, HCO₃⁻ levels had a steady decline due to the reduction of atmospheric CO₂, as documented in the evolution of the terrestrial atmosphere [31]. This reduction could have selected the enzymes 6.3.4.18 and 5.4.99.18, which transform AIR to N-CAIR and N-CAIR to CAIR, respectively.

Subsequently, once the atmosphere was provided with oxygen, the emergence of mitochondria and eukaryotic cells was possible. In this regard, HCO₃⁻ is one of the main products of mitochondrial respiration, which follows a pH-dependent conversion to CO₂, converting an impermeable anion into a gas that can diffuse through membranes [34]. The transformation of HCO₃⁻ accumulated in cells with mitochondrial activity towards CO₂ may have resulted in a more efficient regulation of intracellular pH in eukaryotic cells, and in parallel, to the use of CO₂ as a carbon source. All these processes might have favored multicellularity, because cells with nutrients availability and high mitochondrial activity as the consequence of oxidative respiration could provide a permeable carbon source to other cells with low nutrients availability. Altogether, the taxonomic distribution data, chemical synthesis information, and primitive ocean simulations data, support the divergence of the enzyme 5.4.99.18 to 4.1.1.21 by selecting a CO₂ binding site. Although the selection of a CO₂ binding site in metabolism is not exclusive to this protein family, the selection of

a CO₂ binding site in these enzymes could favor the development of multicellularity in eukaryotic cells.

The last two steps in the synthesis of IMP correspond to the transformation of 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) and 5-formamidoimidazole-4-carboxamide ribotide (FAICAR). In the first step (2.1.2.3), AICAR formyltransferase activity is required, whereas for the second step (3.5.4.10), IMP cyclohydrolase activity is required. The gene for the bifunctional folate-dependent enzyme PurH, which is widely distributed in *Bacteria* and *Eukarya* and partially in *Archaea*, encodes both activities [35]. PurH exhibits two independent catalytic sites, in which each half of the enzyme catalyzes an independent reaction. The transformation to IMP by the bifunctional enzyme PurH may have been posterior to the emergence of the folate synthesis pathway (Figure 1, gold box). In addition, both catalytic reactions performed by PurH could have emerged in *Archaea* to be carried out by different proteins that achieve each step independently. In this regard, the first reaction is carried out by the formate-dependent PurP enzyme, whereas the second one is achieved by the IMP cyclohydrolase PurO (Figure 1, magenta). In both cases, the catalytic mechanism is very similar to that of its counterpart PurH, and they are exclusively distributed in *Archaea*; however, these two enzymes do not maintain relationships at the sequence or structural level with PurH [9,35]. In summary, all these results suggest that PurP and PurO emerged posterior to archaeal divergence.

Alternatively, the archaeal ancestor may have brought the formate (PurT) and folate-dependent (PurN-PurH) families, as seen in *Halobacteria*. The emergence of PurP in *Archaea* was probably related to the availability of formate in the environment, which agrees with the fact that most of the *Archaea* clades are methylophiles. In turn, PurP may have duplicated from families associated with the LCA [carbamoyl-phosphate synthase, 6.3.5.5; phosphoribosylamine-glycine ligase, 6.3.4.13; 5-(carboxyamino) imidazole ribonucleotide synthase, 6.3.4.18; phosphoribosylglycinamide formyltransferase, 2.1.2.2] belonging to nucleotide metabolism, partially replacing PurH in some archaeal clades. For instance, PurH was identified in some archaeal clades, such as *Halobacteria*, *Thermoplasmata*, and *Methanomicrobia* among others, including co-occurrence with PurP, suggesting that their activities are complementary in these organisms.

In the case of PurO, we did not find homology relationships with any other enzymatic family associated with nucleotide biosynthesis. In addition, we found a correlation between the presence of PurP and PurO; however, in *Archaea* some clades do not exhibit either PurH or PurO, which suggests that alternative enzymes performing the PurO function are probably present and remain to be described.

Purine salvage pathway

Ribonucleotide adenine salvage pathway

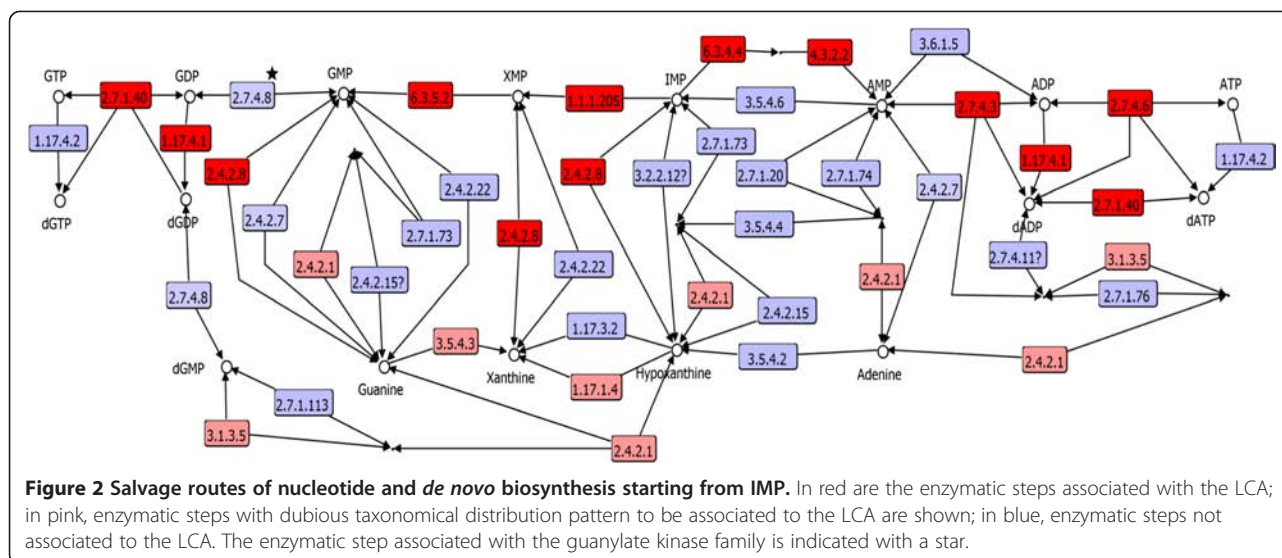
In the purine salvage pathway, diverse pathways to generate adenine and guanine ribonucleotides and deoxynucleotides were identified. In this regard, the first pathway starts from hypoxanthine, via hypoxanthine-guanine phosphoribosyltransferase (PRTase) (2.4.2.8) (Figure 2), an enzyme with broad specificity and that is universally distributed in the three cellular domains. This enzyme achieves the biochemical transformation to IMP, starting from hypoxanthine and PRPP, for subsequent transformation to the ribonucleotide of adenine through the *de novo* adenylyl-succinate pathway, which includes enzymes widely distributed in the three cellular domains (6.3.4.4, adenylosuccinate synthetase; 4.3.2.2, adenylosuccinate lyase; 2.7.4.3, adenylyl kinase; 2.7.4.6, nucleoside diphosphate kinase; 2.7.1.40, pyruvate kinase), suggesting its presence in the LCA (Figure 2). For the transformation of alpha-deoxyribonucleotide, there are two routes, the pathway associated with the ribonucleoside-diphosphate reductase beta-subunit (small subunit) (1.17.4.1), the ribonucleoside-diphosphate reductase alpha-subunit (large chain family), and the 1.17.4.2 pathway, which contains adenosylcobalamin-dependent ribonucleoside-triphosphate reductase, and the anaerobic ribonucleoside-triphosphate reductase complex (NrdD and NrdG). The first of these pathways, for the formation of deoxyribonucleotides, is widely distributed in the three cellular domains, suggesting its presence in the LCA. Finally, the enzymes of the 1.17.4.2 pathway are partially distributed in *Bacteria* and *Archaea* but not in *Eukarya*, which makes it difficult to determine its presence in the LCA.

Alternatively, it is also possible to synthesize adenine deoxynucleotides and ribonucleotides. This pathway conserves the last three catalytic steps (2.7.4.3, 2.7.4.6 and 2.7.1.40) from the pathway previously described, *i.e.*, only the first step has been added, which is catalyzed by adenine phosphoribosyltransferase (a PRTase) (2.4.2.7) (Figure 2), an enzyme widely distributed in *Bacteria* and *Eukarya* and that starts from the adenine salvage pathway.

This enzyme may be specialized for the adenine substrate, posterior to the divergence of the LCA, from a broad-specificity ancestor, similar to the HPRTases (2.4.2.8).

Route of salvage of guanine ribonucleotides

Starting from the first salvaging route, previously described, which starts with hypoxanthine-guanine phosphoribosyltransferase (2.4.2.8), it is also possible to synthesize guanine ribonucleotide, adding to the enzymatic families widely distributed in the three cellular domains: inosine-5'-monophosphate dehydrogenase 1 (1.1.1.205) and GMP synthase (glutamine-hydrolyzing) subunit A and GMP synthase (glutamine-hydrolyzing) subunit B (6.3.5.2)



(Figure 2). Because the hypoxanthine-guanine phosphoribosyltransferase (2.4.2.8) also exhibits specificity for guanine, it is possible to synthesize (ribonucleotide monophosphate guanine) guanosine 5'-monophosphate (GMP) in one step, starting from the guanine salvage pathway with PRPP; however, although the subsequent step with 2.7.4.8, performed by the guanylate kinase family, for transforming GMP to GDP is widely distributed in *Eukarya* and *Bacteria*, in *Archaea* only the reaction but not the enzyme has been identified (Figure 2, star) [36]. These data limit the possibility of making genomic and evolutionary comparisons between *Bacteria* and *Eukarya* to guarantee the possible presence of guanine deoxyribonucleotides in the LCA. However, subsequent steps for the synthesis of GDP to GTP are carried out with the nucleoside diphosphate kinase (2.7.4.6) and pyruvate kinase (2.7.1.40); both of these enzymes are widely distributed in the three cellular domains, and the step of GTP conversion to dGTP for the widely distributed enzyme ribonucleoside diphosphate reductase (1.17.4.1) suggests the presence of GTP and dGTP in the LCA.

Therefore, although the gene sequence associated with the guanylate kinase function in *Archaea* has not yet been identified, we suggest that this protein could exhibit a common origin with its counterpart in *Bacteria* and *Eukarya*, because most of the nucleotide kinases belong to the P-loop-containing nucleoside triphosphate hydrolases family. This family shows a broad specificity in recognition of ribonucleotides and deoxyribonucleotides [37]. In this regard, it has been reported that only two mutations are enough to introduce the adenylate kinase activity into guanylate kinase [38], suggesting a masking of guanylate kinase assignment by adenylate kinase. In addition, it is possible that an enzyme with broad specificity similar to adenylate kinase (Figure 3, star) recognizes NMP and

converts it into NDP; thus, its specialization toward the guanylate kinase function occurred posterior to the divergence of the three cellular domains.

At present, members of the adenylate kinase family and in particular from the AK6 subfamily have been described as proteins that exhibit the ability to transform different NMPs to their corresponding NDPs, specifically, the AK6 subfamily transforms AMP, dAMP, CMP, dCMP, IMP, and GMP to their corresponding NDPs and dNDPs [39,40]. This finding also suggests that an ancestral enzyme with broad specificity for the synthesis of ribonucleotides and deoxynucleotides could have been utilized in the LCA. It is also interesting that this enzyme could have used IMP as a substrate, since this nucleotide could have had an active role in the evolution of nucleic acids and the genetic code [41]. Interestingly, our results show that *Archaea* contain an adenylate kinase from the AK6-like subfamily, similar to *Eukarya* and *Bacteria*, suggesting that the function of guanylate kinase in archeal organisms, which has not yet been identified, could be carried out at least in part by enzymes of this subfamily.

Integrating the evolutionary analysis of the salvage and *de novo* purine pathways

The taxonomic distribution of enzymes associated with purine metabolism shows that segments of the *de novo* and salvage pathways were complementary and critical to the availability of nucleic acids before the divergence of the three cellular domains. This finding correlates with the chemical synthesis under prebiotic conditions for nitrogenous base precursors for the purine salvage pathway, as previously described [10,13,42,43] (Figure 3, blue circles). The findings of those previous studies agree with the fact that these routes were dependent on PRPP biosynthesis, since it has been argued that PRPP,

bifurcation of the *de novo* IMP pathway and also with AICA in backflow as a seed substrate (Figure 3, following the direction of the dashed arrows). Additionally, the thiamine pathway may have been also fed from a third source by means of histidine metabolism (Figure 3, box with bars); the connection to purine biosynthesis results from an enzymatic step catalyzed by imidazole glycerol phosphate (IGP) synthase, which transforms *N*-(5-phosphoribosyl)-formimino-5-aminoimidazol-4-carboxamide ribonucleotide (PREAR) into AICAR, which is then recycled into the *de novo* purine biosynthetic pathway, and imidazole-glycerol 3-phosphate, which leads to histidine. Interestingly, previous works have also suggested that the histidine biosynthetic route is ancient and related to the emergence of life [18]. Our proposal of an ancestral branch related to thiamine synthesis is consistent with the catalytic functions that have been proposed for this molecule in the early evolution of life, as suggested by its essential catalytic role in most of organisms and its requirement at several central points of anabolic and catabolic intermediary metabolism [44], as well as in semienzymatic pathways that may precede the actuals [45]. It is interesting this branch in the early emergence of life and before the constitution of the LCA could have been fed the thiamine metabolism in a semienzymatic way, as the AICAR-to-AIR transformation could occur in a facile, nonenzymatic chemical synthesis pathway [46].

In this regard, molecules of thiamine or its derivatives bind to the mRNA in the absence of cofactors or proteins in the three domains of life, forming a complex that sequesters the ribosome binding site and suggesting the existence of an ancestral form of riboswitches, which have been implicated in regulatory mechanisms [47]. Additionally, thiamine could have interacted with the RNA, leading to catalytically versatile ribozymes related to the RNA world, due to its catalytic and RNA binding capabilities [44].

Finally, we suggest that the guanylate kinase function could have been carried out in the LCA by an ancestral enzyme with broad specificity to nucleoside monophosphates (NMPs), similar to those in the AK6 adenylate kinase subfamily. Its specialization to guanylate kinase could have been posterior to archaeal divergence. In parallel, this ancestral enzyme may have provided the nucleotide inosine (Figure 3, pink circle), which has been suggested to play an active role in a rudimentary stage of the genetic code [41]. The ITP can compete with or replace the ATP and GTP binding sites in diverse proteins, such as RNA polymerase [48,49], that may have maintained their affinity and specificity for ITP as a remnant of its ancestral role.

In addition, inosine maintains a strong structural similarity to guanine [50], and it may even have the same function for guanine pairing to cytosine in the codons. Currently, inosine has been found in the third position

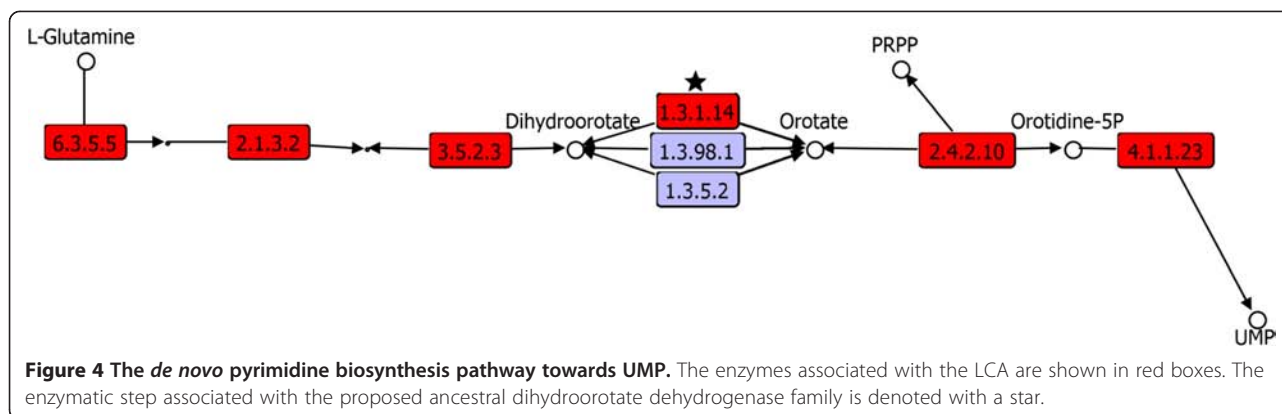
of anticodons, pairing with codons at bases U, C or A and thereby decreasing the need for the 61 tRNAs for each codon. In this regard, it has been suggested that inosine may have been produced by adenosine deamination or even in RNA-mediated catalysis through an early stage of emergence of the genetic code, and it was excluded in nucleic acids when the canonical Watson-Crick pairing evolved to avoid ambiguous rules in replication [41]. Posterior to the divergence of *Archaea* from the LCA, the specialization and divergence to guanylate kinase of members with broad specificity of the ancestral AK6 subfamily type could have increased the availability of guanine, favoring the replacement of inosine for guanine.

Evolution of pyrimidine metabolism

De novo pyrimidine biosynthesis

Based on taxonomical distributions, we evaluated the enzymes associated with *de novo* pyrimidine biosynthesis: carbamoyl-phosphate synthase large chain, carbamoyl-phosphate synthase small chain, 6.3.5.5; aspartate carbamoyltransferase 1, 2.1.3.2; dihydroorotase, DHOase family, 3.5.2.3; dihydroorotate dehydrogenase B (NAD⁺), catalytic subunit, dihydroorotate oxidase B, electron transfer subunit, 1.3.1.14; dihydroorotate dehydrogenase (quinone), 1.3.5.2; dihydroorotate dehydrogenase A (fumarate), 1.3.98.1; orotate phosphoribosyltransferase 1, PyrE1, 2.4.2.10; orotidine 5'-phosphate decarboxylase, 4.1.1.23; cytidylate kinase, 2.7.4.14; nucleoside diphosphate kinase, 2.7.4.6 (Figure 4). These enzymes form the branch of UTP *de novo* biosynthesis starting from L-glutamine. This entire branch may have occurred in the universal ancestor, based on the wide taxonomical distribution of the enzymes that compose it (Figure 4, red boxes). One of the most interesting steps of this pathway is the conversion of dihydroorotate to orotate, which is carried out by enzymes classified into two families of dihydroorotate group dehydrogenases (DHODs), according to the terminal electron acceptor and relationships at the sequence level [51]. In this regard, family 1 uses soluble electron acceptors. This family is widely distributed in gram-positive *Bacteria*, *Archaea* and in some unicellular eukaryotic organisms. In turn, this family is subdivided into DHODA (1.3.98.1) and DHODB (1.3.1.14), which are homodimeric and heterotetrameric, respectively. DHODA uses fumarate as the electron acceptor, whereas DHODB uses NAD⁺ [52]. Members of family 2 (1.3.5.2) are linked to the cell membrane and use quinones from the respiratory chain as electron acceptors [53]. These enzymes are mainly found in most eukaryotic organisms and gram-negative bacteria, in agreement with our taxonomical distribution results.

In spite of these differences, both DHODs belong to the FMN-linked oxidoreductases superfamily, suggesting a common ancestor. Probably, this ancestral enzyme was similar in functional terms to members of the family 2



(Figure 4, star), according to its electron acceptor molecules, which have been previously described as one of the most abundant in extraterrestrial environments, and also that could have been delivered to the Earth around 4 billions years ago, making possible the prebiotic conditions needed for the emergence of life. Indeed, quinones have been found in meteorites in considerable amounts and also have been synthesized in a cloud chamber simulation with good yields [54,55]. Additionally, quinones are spontaneously partitioned into model membrane systems, representing an evolutionary advantage to early organisms by providing some protection against UV radiation in the early Earth environment [56], and were posteriorly exploited for their capacity to pump protons across membrane bilayers [57]. In a posterior step, changes associated with the cell membrane and cell wall could lead to divergence of the DHODs from family 2 to family 1, via incorporation of soluble electron acceptors. Surprisingly, our proposal for DHOD family divergence is consistent with previous reports describing a transition from gram-negative to gram-positive bacteria [58,59] and to drastic changes associated with the chemical constituents of the cell membrane in *Archaea*, such as glycerol stereochemistry [60], posterior to the divergence of the LCA. Furthermore, subdivision of family 1 might have occurred in the direction of DHODB to DHODA, since most of these enzymes have been found in gram-positive bacteria that have adapted to parasitic or symbiotic relationships, as shown by our taxonomical distribution results. The DHODA uses fumarate as a terminal electron acceptor, which in turn is used instead of oxygen as a terminal electron acceptor for succinate production, one of the essential processes controlling redox homeostasis for many parasites living under anaerobic conditions [53].

A wide taxonomic distribution of the enzyme CTP synthase (6.3.4.2), which converts UTP to CTP, was also found in the three cellular domains, in addition to the enzyme for the subsequent steps, nucleoside diphosphate kinase (2.7.4.6), thioredoxin reductase NTRA (1.8.1.9) and

ribonucleotide reductase (1.17.4.1), giving rise to deoxy-CTP. Based on the distributions of these enzymes, the presence of UTP, CTP and deoxy-CTP in the LCA is suggested (Figure 5, red boxes).

Concerning the transformation to deoxyuridine, two main routes have been described. One pathway starts from deoxy-CDP, whereas the second one starts from deoxy-CMP and which can be extended to transforming deoxyuridine. The route of deoxy-CDP requires two enzymatic steps: the first one requires the nucleoside diphosphate kinase (2.7.4.6), which produces dUTP, which is subsequently transformed by the inosine triphosphate pyrophosphatase 1 (nucleoside-triphosphate pyrophosphatase 1; 3.6.1.19) to dUMP. These two enzymes were identified as universally distributed in the three cellular domains; therefore, they can be associated with the LCA (Figure 5, red boxes). The second pathway converts dCMP to dUMP via cytidine/deoxycytidylate deaminase, which is associated with the catalytic activity of 3.5.4.12, whose members are partially distributed among the three cellular domains. Finally, a third pathway, which starts from dCTP and is catalyzed by deoxycytidine triphosphate deaminase (3.5.4.13), is absent in eukaryotes and partially distributed in *Bacteria* and *Archaea*. The assignment of the second and third pathways toward deoxyuridine transformation was within the LCA, based on the complex evolutionary history of their enzymes.

Finally, the transformation of deoxy-UMP to deoxy-TMP can be carried out through two folate-dependent enzymes, thymidylate synthase ThyX (2.1.1.148) and thymidylate synthase ThyA (2.1.1.45) [9] (Figure 5 in gold boxes). These enzymes are not homologous, suggesting an independent evolutionary origin. Both enzymes promote methylation by using the 5,10-methylenetetrahydrofolate ($\text{CH}_2\text{-H}_4$ folate) as a carbon donor. ThyA also uses $\text{CH}_2\text{-H}_4$ folate to produce dihydrofolate ($\text{H}_2\text{-folate}$). In contrast, ThyX uses flavin adenine dinucleotide (FAD) and NAD(P)H as cofactors to form reduced tetrahydrofolate ($\text{H}_4\text{-folate}$) [61]. ThyX is partially distributed in *Bacteria*

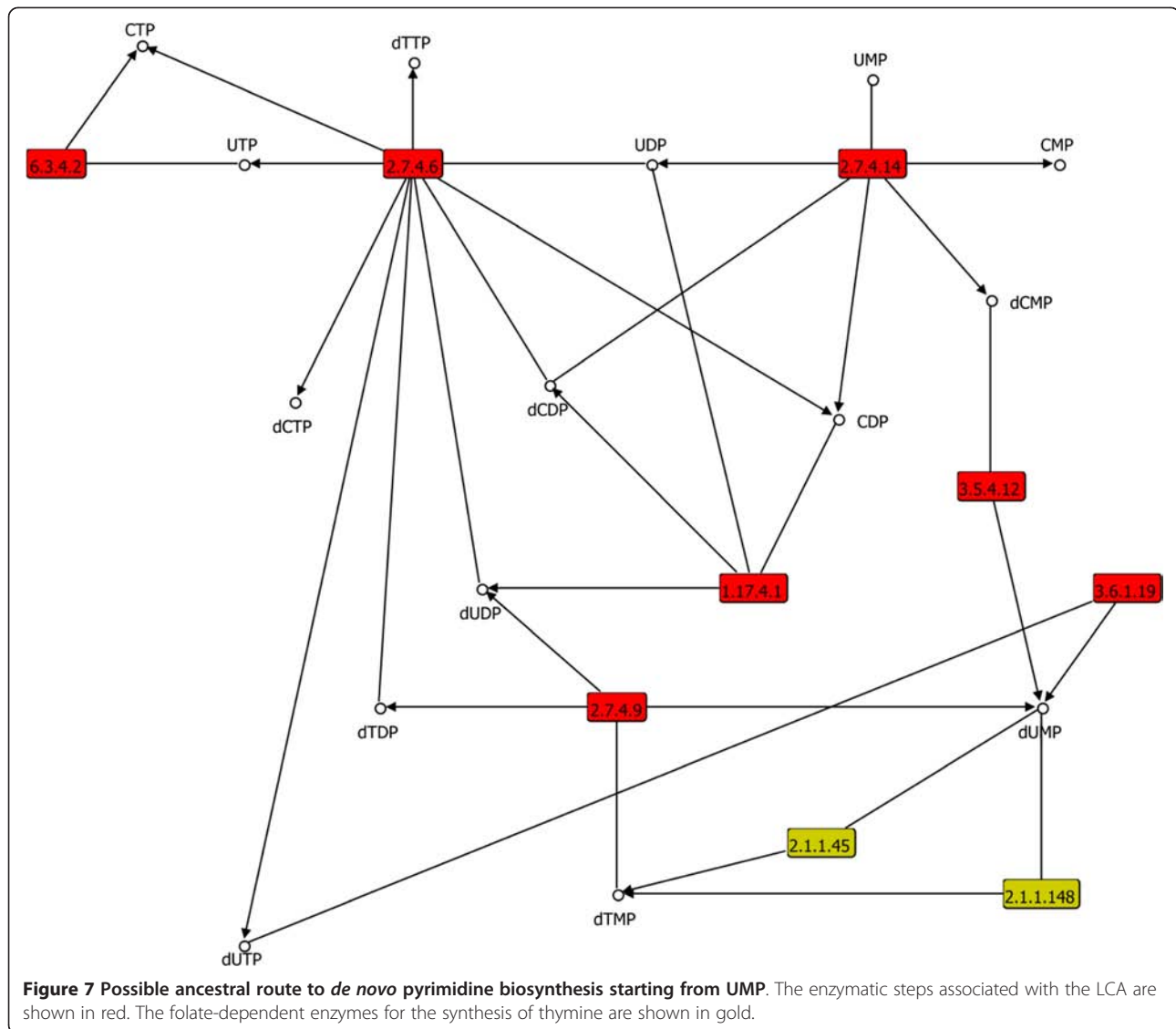


Figure 7 Possible ancestral route to *de novo* pyrimidine biosynthesis starting from UMP. The enzymatic steps associated with the LCA are shown in red. The folate-dependent enzymes for the synthesis of thymine are shown in gold.

correlated with the transition from gram-negative to gram-positive bacteria and was affected by similar evolutionary pressures. The evolutionary pressures could be associated with a variety of environmental changes, such as increased atmospheric oxygen levels, temperature or changes from water to soil habitats. These changes could lead to modifications in the plasma membrane's chemical properties, resulting in the divergence of family 2 enzymes (1.3.5.2) linked to the cell membrane, which uses quinones from the respiratory chain as electron acceptors, to the family 1 enzymes, which incorporate soluble electron acceptors. This transition from gram-negative to gram-positive was also suggested by Cavalier-Smith [59,60]; altogether, with our data regarding the taxonomic distribution showing that family 1 is ubiquitous in gram-positive bacteria and archaea, we therefore suggest

close evolutionary relationships between archaea and gram-positive bacteria.

Although gram-negative bacteria and eukarya contain the subfamily 2 of DHODS, it is not possible to deduce an evolutionary relationship, as we previously identified between gram-positive bacteria and archaea, because members of this subfamily in eukarya are linked to the mitochondrial membrane. In this regard, the mitochondrial acquisition in *Eukarya* has been described as a probable lateral gene transfer event, as described in the endosymbiont theory [64]. Therefore, our results support the notion of phagocytosis of a gram-negative bacterium by a protoeukaryotic cell, with posterior specialization to mitochondria.

Additionally, *de novo* synthesis of cytosine ribonucleotides and deoxyribonucleotides was probably associated

with the LCA. In *de novo* biosynthesis of thymine, the folate-dependent enzymes ThyA and ThyX can catalyze the transformation of this metabolite independently, suggesting that this pathway appeared posterior to the emergence of folate biosynthesis. For enzymes present in the thymine salvage pathway, it was not possible to determine their presence in the LCA.

Conclusion

The analysis presented here is based on multiple complete genomes belonging to organisms from the three cellular domains, along with current biochemical knowledge. These analyses allowed us to identify issues related to the origin and evolution of nucleotide metabolism. One of our main findings is that we were able to assess the ancestry of some segments of the purine salvage and the *de novo* pathways, which could be complementary and closely related to the LCA of the three cellular domains. Additionally, it was found that a large part of the *de novo* purine branch is widely distributed in *Bacteria*, *Archaea* and *Eukarya*, primarily towards *de novo* biosynthesis of IMP (a key precursor of purines). This branch may have been associated, in the early stages of cell evolution, with the metabolism of thiamine (vitamin B₁) and posteriorly was complemented by the addition of two new enzymatic steps to complete the IMP biosynthesis pathway by means of the folate-dependent PurH enzyme, giving rise to the modern *de novo* synthesis of purines.

The ancestry and divergence of enzymes associated with these routes provide clues to the environmental changes in the early stages of the emergence of life. Such is the case with the divergence of the enzyme phosphoribosylaminoimidazole carboxylase (4.1.1.21) from N⁵-carboxyaminoimidazole ribonucleotide mutase (5.4.99.18); the divergence of these two enzymes supports the hypothesis of the origin of life in primitive seas with high levels of HCO₃⁻. Once the atmosphere was provided with oxygen and the first eukaryotic organisms with mitochondria emerged, the enzyme 5.4.99.18 diverged to 4.1.1.21 through acquisition of a CO₂ binding site. This divergence process, among other metabolic changes, may have facilitated the emergence of the first eukaryotic multicellular organisms.

In the case of pyrimidines, it was possible for us to infer that the LCA synthesized uracil ribonucleotides, by both the *de novo* and salvage pathways, suggesting that this ribonucleotide could have been involved in a great number of enzymatic functions and/or regulation as a remnant of the RNA world. Additionally, it was possible to associate the synthesis of cytosine and uracil deoxyribonucleotides in the LCA, and once folate biosynthesis was possible, the thymine deoxyribonucleotides emerged due to their enzymatic dependence on the folate precursor. Before the

emergence of folate biosynthesis, some variants of these three bases, including cytosine or cytosine-methylated derivatives, based on its similarity to thymine, could have played a role similar to actual thymine in DNA. This inference is supported by the fact that several analogs of cytosine and uracil partially integrate into DNA as replacements for thymine [65-67]. Thus, thymine does not have an active role in the nucleotide transcript and it is more likely to be replaced. This argument is consistent with the evolution of bacterial strains [68], which shows that strains can be generated with the ability to incorporate derivatives of uracil (chloro-uracil) by replacing up to 98% of thymine and maintain cell viability.

In relation to purine synthesis, our results revealed that the biosynthesis of adenine could have been carried out in the LCA by the adenylate kinase. In the case of guanine biosynthesis, we found a complex evolutionary history; for instance, its synthesis has been detected in archaeal organisms, although its gene sequence has not been identified [36]. The fact that we did not find the guanylate kinase bacterial type in *Archaea* suggests three possible scenarios: a) a high sequence divergence between these proteins, b) a masking function for some other enzyme, and/or c) a gene loss or gene nonorthologous displacement. As discussed above, it is likely that the guanylate kinase function has been masked by the adenylate kinase AK6 subfamily type (AK6). This enzyme is homologous to the canonical guanylate kinase of *Eukarya* and *Bacteria* and is probably closer to the possible ancestral enzyme present in the LCA, with promiscuity in the synthesis of NMPs and dNMPs to their respective NDPs and dNDPs. In addition, we found the presence of this multispecific adenylate kinase AK6 subfamily type in *Archaea*, suggesting that this could have been acquired from the LCA with no major posterior changes. In contrast, the guanylate kinase of *Bacteria* and *Eukarya* may have diverged to a greater extent from the adenylate kinase family subsequent to archaeal divergence. The structural similarity of inosine and guanine, together with the remnant affinity of inosine in several proteins such as RNA polymerase, in addition to the taxonomical distribution data that showed a possible LCA route for the synthesis of ITP (Figure 3, magenta circle) by means of a broad-specificity ancestral enzyme similar to the adenylate kinase Ak6 subfamily type, suggest that this base could have played an important role in early cell evolution, as has been previously proposed [41]. The ancestral enzyme might have provided both IDP and GDP for subsequent processing to NTPs, with both playing similar informational process roles in the genetic code based on their structural similarities.

The subsequent specialization of guanylate kinase may have facilitated greater availability of guanine nucleotide to replace inosine, thus avoiding ambiguous rules in DNA

replication that would have been achieved with this consolidation stage of the genetic code.

Interestingly, the divergence of the DHOD families in the LCA suggests transitions associated with changes in the cell wall and cell membrane, supporting an order of divergence from cell walls of gram-negative-like organisms and membranes similar to *Eukarya-Bacteria*, towards gram-positive cell wall, and/or membranes similar to Archaea. Since the plasmatic membrane is considered a matter of vertical inheritance, we suggest that the divergence of the family of DHODs can be associated with the divergence of *Archaea* and gram-positive bacteria.

Methods

Profiles (RPS-Blast)

In order to select the enzymatic numbers and their corresponding enzymes belonging to nucleotide metabolism, the KEGG [2] and MetaCyc [69] databases were exhaustively explored. In total, 120 enzymatic numbers and their corresponding enzymes were collected. (Additional file 5: Table S3 and Additional file 6: Table S4). In a second step, RPS-Blast profiles were used to search for the occurrence of members of enzyme families in complete genomes. These profiles were extracted from PRIAM, the specialized database for detection of enzymatic sequences [70]. This database encompasses characteristics representative of alignments manually annotated, including members of a particular enzyme family, according to the Enzyme DB [71]. In addition, enzymatic steps can be associated with more than one profile, such as protein complexes, or families of nonenzymatic homologous proteins (analogous enzymes) or steps performed by members of the same families (paralogs). Altogether, profiles were curated manually based on functional annotated domains according to Enzyme DB [71].

Enzymatic function

For the analysis of enzymatic functions, the best hit for each sequence, with an E-value of $\leq 10^{-10}$ and coverage of $\geq 55\%$ in relation to the profile, was considered. Similar criteria have been previously described for enzymatic annotation of complete genomes [72].

Taxonomical distributions

In total, 151 sequence profiles associated with 120 enzymatic reactions related to nucleotide metabolism were evaluated in 2,044 complete genomes. The evaluation was based on profile comparisons, using RPS-Blast against the 2,044 complete genomes. Organisms classified as obligate parasites or those organisms with a reduced genome, *i.e.*, less than 1,000 genes, were not considered in this study, with the aim of excluding a possible bias associated with massive gene loss, as previously described [1,8], leaving us with a total of 1,606 of the 2,044 complete genomes. In

order to exclude redundancy in the genomes analyzed, we clustered the organisms based on their taxonomical classifications, in order to obtain a normalized measure of the taxonomical distribution of enzymes according to the following steps. In the first step, we obtained the average presence/absence of enzymes per genus, which in turn was used for the second step, to calculate the average presence/absence of enzymes per Clade. Clades corresponded to the taxonomical categories from the Joint Genome Institute's Integrated Microbial Genomics. Finally, we consider enzymes as widely distributed as those present in more than 50 percent of the clades of the same cellular domain. Additional file 1: Figure S1 and Additional file 2: Figure S2.

Identification of structural domains and families

Structural domains of proteins and families were identified using the models deposited in the Superfamily database [73]. In brief, this database contains Hidden Markov models (HMMs) for each superfamily, which are classified according to their structural domains, which in turn are based on the classifications of SCOP [74]; all members of the same superfamily have a common evolutionary origin. Consensus sequences were derived from the alignments (from which profiles were constructed) and used to identify their corresponding superfamilies, using default parameters.

Additional files

Additional file 1: Figure S1. Average taxonomic distribution of purine metabolism enzymes distributed across the three domains of life. The taxonomic distribution for enzymes catalyzing the purine metabolism (vertical labels) was computed by searching for their ortholog distribution across diverse taxonomic groups (*Archaea*, *Bacteria* and *Eukarya*) (horizontal labels). Enzymes are sorted in terms of their E.C. number. Some identical E.C. numbers and superfamily assignment are complementary according to PRIAM db. The "S" notation defines protein subunits; "AN" defines ANalogous enzymes, those defined as enzymes with the same E.C. number and the different superfamily classification, suggesting different evolutionary origin. The colors are in rainbow scale, where dark blue are those enzymes with low average (0) whereas red denotes high average values (1).

Additional file 2: Figure S2. Average taxonomic distribution of pyrimidine metabolism enzymes. Labels and colors are as in Additional file 1: Figure S1. In asterisks are indicated those enzymes common to purine and pyrimidine metabolisms.

Additional file 3: Table S1. Matrix used to construct the average taxonomic distribution. Purine metabolism. Column 1 shows the E.C. numbers; columns 2 to 76 correspond to the taxonomical divisions analyzed here. Each cell includes normalized values as described in material and methods. (XLSX 35 kb).

Additional file 4: Table S2. Matrix used to construct the average taxonomic distribution. Pyrimidine metabolism. Column 1 shows the E.C. numbers; columns 2 to 76 correspond to the taxonomical divisions analyzed here. Each cell includes normalized values as described in material and methods.

Additional file 5: Table S3. Enzymes analyzed. Purine metabolism (Map 00230). Column 1 denotes the E.C. Number; column 2, the accepted name;

column 3, reactions associated to each E. C. number in the purine metabolism; and column 4, the identification code on KEGG database.

Additional file 6: Table S4. Enzymes analyzed. Pyrimidine metabolism (Map 00240). Column 1 denotes the E.C. Number; column 2, the accepted name; column 3, reactions associated to each E. C. number in the pyrimidine metabolism; and column 4, the identification code on KEGG database.

Competing interest

The authors declare that they have no competing interests.

Author's contribution

DA-M retrieved all enzymatic repertoires, performed the all sequence analysis and interpreted the results. LS interpreted the results; and EP-R interpreted the results. DA-M and EP-R drafted and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Georgina Hernandez-Montes and Jessica Brambila-Tapia for their critical reading of the manuscript. DA-M acknowledges the Programa de Maestría y Doctorado en Ciencias Bioquímicas at Instituto de Biotecnología-UNAM and his support by a PhD fellowship (35083) from CONACYT and (IN-204714) DGAPA. EP-R was supported by a grant (IN-204714) from DGAPA and (155116) from CONACYT. We also want to thank Juan Manuel Hurtado Ramirez by his technical support.

Author details

¹Departamento de Ingeniería Celular y Biotecnología, Instituto de Biotecnología, UNAM Av. Universidad 2001, Cuernavaca, Morelos CP 62210, México.

²Present address: Unidad Multidisciplinaria de Docencia e Investigación, Sisal Facultad de Ciencias, UNAM, Sisal, Yucatán, México.

Received: 25 March 2014 Accepted: 15 September 2014

Published: 17 September 2014

References

- Hernandez-Montes G, Diaz-Mejia JJ, Perez-Rueda E, Segovia L: **The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution.** *Genome Biol* 2008, **9**(6):R95.
- Wixon J, Kell D: **The Kyoto encyclopedia of genes and genomes-KEGG.** *Yeast* 2000, **17**(1):48-55.
- Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet* 2004, **5**(2):101-113.
- Diaz-Mejia JJ, Perez-Rueda E, Segovia L: **A network perspective on the evolution of metabolism by gene duplication.** *Genome Biol* 2007, **8**(2):R26.
- Horowitz NH: **On the evolution of biochemical syntheses.** *Proc Natl Acad Sci U S A* 1945, **31**(6):153.
- Jensen RA: **Enzyme recruitment in evolution of new function.** *Annu Rev Microbiol* 1976, **30**:409-425.
- Lopez de la Osa J, Bateman DA, Ho S, Gonzalez C, Chakrabarty A, Laurents DV: **Getting specificity from simplicity in putative proteins from the prebiotic earth.** *Proc Natl Acad Sci U S A* 2007, **104**(38):14941-14946.
- Delaye L, Becerra A, Lazcano A: **The last common ancestor: what's in a name?** *Orig Life Evol Biosph* 2005, **35**(6):537-554.
- Brown AM, Hoopes SL, White RH, Sarisky CA: **Purine biosynthesis in archaea: variations on a theme.** *Biol Direct* 2011, **6**:63.
- Becerra A, Lazcano A: **The role of gene duplication in the evolution of purine nucleotide salvage pathways.** *Orig Life Evol Biosph* 1998, **28**(4-6):539-553.
- Caetano-Anolles G, Yafremava LS, Gee H, Caetano-Anolles D, Kim HS, Mittenthal JE: **The origin and evolution of modern metabolism.** *Int J Biochem Cell Biol* 2009, **41**(2):285-297.
- Oro J, Yamrom T, Cortes S: **The Prebiotic Synthesis of Imidazoles and Their Catalytic Role in the Synthesis of Oligonucleotides.** *Fed Proc* 1983, **42**(7):1926-1926.
- Basile B, Lazcano A, Oro J: **Prebiotic syntheses of purines and pyrimidines.** *Adv Space Res* 1984, **4**(12):125-131.
- Meléndez-Hevia E, Waddell TG, Montero F: **Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle.** *J Theor Biol* 1994, **166**(2):201-220.
- Morowitz H, Peterson E, Chang S: **The Synthesis of Glutamic-Acid in the Absence of Enzymes - Implications for Biogenesis.** *Orig Life Evol Biosph* 1995, **25**(4):395-399.
- Waddell TG, Eilders LL, Patel BP, Sims M: **Prebiotic methylation and the evolution of methyl transfer reactions in living cells.** *Orig Life Evol Biosph* 2000, **30**(6):539-548.
- Peregrín-Alvarez JM, Sanford C, Parkinson J: **The conservation and evolutionary modularity of metabolism.** *Genome Biol* 2009, **10**(6):R63.
- Fani R, Fondi M: **Origin and evolution of metabolic pathways.** *Phys Life Rev* 2009, **6**(1):23-52.
- Kyrpides N, Overbeek R, Ouzounis C: **Universal protein families and the functional content of the last universal common ancestor.** *J Mol Evol* 1999, **49**(4):413-423.
- Koonin EV: **Comparative genomics, minimal gene-sets and the last universal common ancestor.** *Nat Rev Microbiol* 2003, **1**(2):127-136.
- Zrenner R, Stitt M, Sonnewald U, Boldt R: **Pyrimidine and purine biosynthesis and degradation in plants.** *Annu Rev Plant Biol* 2006, **57**:805-836.
- Vantrump JE, Miller SL: **Carbon-Monoxide on Primitive Earth.** *Earth Planet Sc Lett* 1973, **20**(1):145-150.
- Smirnov A, Schoonen MA: **Evaluating experimental artifacts in hydrothermal prebiotic synthesis experiments.** *Orig Life Evol Biosph* 2003, **33**(2):117-127.
- Martin W, Russell MJ: **On the origin of biochemistry at an alkaline hydrothermal vent.** *Philos Trans R Soc Lond B Biol Sci* 2007, **362**(1486):1887-1925.
- Wallace KB, Starkov AA: **Mitochondrial targets of drug toxicity.** *Annu Rev Pharmacol Toxicol* 2000, **40**:353-388.
- Haritos VS, Dojchinov G: **Cytochrome c oxidase inhibition in the rice weevil *Sitophilus oryzae* (L.) by formate, the toxic metabolite of volatile alkyl formates.** *Comp Biochem Physiol C Toxicol Pharmacol* 2003, **136**(2):135-143.
- Kappock TJ, Ealick SE, Stubbe J: **Modular evolution of the purine biosynthetic pathway.** *Curr Opin Chem Biol* 2000, **4**(5):567-572.
- Zhang Y, Morar M, Ealick SE: **Structural biology of the purine biosynthetic pathway.** *Cell Mol Life Sci* 2008, **65**(23):3699-3724.
- Tribunskikh IA, Alenin VV, Selivanov SI, Shavva AG, Inge-Vechtomo SG: **Divergence of de novo biosynthesis of inosine-5'-triphosphate.** *Dokl Biochem Biophys* 2005, **400**:65-68.
- Dismukes G, Klimov V, Baranov S, Kozlov YN, DasGupta J, Tyrshkin A: **The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis.** *Proc Natl Acad Sci* 2001, **98**(5):2170-2175.
- Grotzinger JP, Kasting JF: **New constraints on Precambrian ocean composition.** *J Geol* 1993, **101**(2):235-243.
- Meyer E, Kappock TJ, Osuji C, Stubbe J: **Evidence for the direct transfer of the carboxylate of N5-carboxyaminoimidazole ribonucleotide (N5-CAIR) to generate 4-carboxy-5-aminoimidazole ribonucleotide catalyzed by *Escherichia coli* PurE, an N5-CAIR mutase.** *Biochemistry* 1999, **38**(10):3012-3018.
- Cusack NJ, Shaw G, Litchfield GJ: **Purines, pyrimidines, and imidazoles. XXXVI. Carboxylation of some 5-aminoimidazoles and related compounds, including nucleosides and nucleotides, with potassium hydrogen carbonate in aqueous solution.** *J Chem Soc Perkin 1* 1971, **8**:1501-1507.
- Cordat E, Casey JR: **Bicarbonate transport in cell physiology and disease.** *Biochem J* 2009, **417**:423-439.
- Kang YN, Tran A, White RH, Ealick SE: **A novel function for the N-terminal nucleophile hydrolase fold demonstrated by the structure of an archaeal inosine monophosphate cyclohydrolase.** *Biochemistry* 2007, **46**(17):5050-5062.
- Grochowski LL, Censky K, Xu HM, White RH: **A new class of adenylate kinase in methanogens is related to uridylylase kinase.** *Arch Microbiol* 2012, **194**(2):141-145.
- Yan H, Tsai MD: **Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity.** *Adv Enzymol Relat Areas Mol Biol* 1999, **73**:103-134.
- Stolworthy TS, Black ME: **The mouse guanylate kinase double mutant E72Q/D103N is a functional adenylate kinase.** *Protein Eng* 2001, **14**(11):903-909.
- Feng X, Yang RN, Zheng XF, Zhang FY: **Identification of a novel nuclear-localized adenylate kinase 6 from *Arabidopsis thaliana* as an essential stem growth factor.** *Plant Physiol Biochem* 2012, **61**:180-186.
- Ren H, Wang L, Bennett M, Liang Y, Zheng X, Lu F, Li L, Nan J, Luo M, Eriksson S, Zhang C, Su XD: **The crystal structure of human adenylate kinase 6: An adenylate kinase localized to the cell nucleus.** *Proc Natl Acad Sci U S A* 2005, **102**(2):303-308.
- Keegan LP, Leroy A, Sproul D, O'Connell MA: **Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes.** *Genome Biol* 2004, **5**(2):209.

42. Hitchings GH: The purine metabolism of protozoa. *Adv Enzyme Regul* 1982, **20**:375–386.
43. Zubay G: To what extent do biochemical pathways mimic prebiotic pathways. *Chemtracts Biochem Mol Biol* 1993, **4**:317–323.
44. Frank RA, Leeper FJ, Luisi BF: Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cell Mol Life Sci* 2007, **64**(7–8):892–905.
45. Lazcano A, Miller SL: On the origin of metabolic pathways. *J Mol Evol* 1999, **49**(4):424–431.
46. Groziak MP, Bhat B, Leonard NJ: Nonenzymatic synthesis of 5-aminoimidazole ribonucleoside and recognition of its facile rearrangement. *Proc Natl Acad Sci U S A* 1988, **85**(19):7174–7176.
47. Winkler W, Nahvi A, Breaker RR: Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 2002, **419**(6910):952–956.
48. Sakumi K, Abolhassani N, Behmanesh M, Iyama T, Tsuchimoto D, Nakabeppu Y: ITPA protein, an enzyme that eliminates deaminated purine nucleoside triphosphates in cells. *Mutat Res* 2010, **703**(1):43–50.
49. Glasemacher J, Bock AK, Schmid R, Schonheit P: Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. *Eur J Biochem* 1997, **244**(2):561–567.
50. Krepl M, Otyepka M, Banas P, Sponer J: Effect of Guanine to Inosine Substitution on Stability of Canonical DNA and RNA Duplexes: Molecular Dynamics Thermodynamics Integration Study. *J Phys Chem B* 2013, **117**(6):1872–1879.
51. Rowland P, Norager S, Jensen KF, Larsen S: Structure of dihydroorotate dehydrogenase B: electron transfer between two flavin groups bridged by an iron-sulphur cluster. *Structure* 2000, **8**(12):1227–1238.
52. Takashima E, Inaoka DK, Osanai A, Nara T, Odaka M, Aoki T, Inaka K, Harada S, Kita K: Characterization of the dihydroorotate dehydrogenase as a soluble fumarate reductase in *Trypanosoma cruzi*. *Mol Biochem Parasitol* 2002, **122**(2):189–200.
53. Norager S, Jensen KF, Bjornberg O, Larsen S: E-coli dihydroorotate dehydrogenase reveals structural and functional distinctions between different classes of dihydroorotate dehydrogenases. *Structure* 2002, **10**(9):1211–1223.
54. Bernstein MP, Sandford SA, Allamandola LJ, Gillette JS, Clemett SJ, Zare RN: UV irradiation of polycyclic aromatic hydrocarbons in ices: production of alcohols, quinones, and ethers. *Science* 1999, **283**(5405):1135–1138.
55. Deamer D, Dworkin JP, Sandford SA, Bernstein MP, Allamandola LJ: The first cell membranes. *Astrobiology* 2002, **2**(4):371–381.
56. Wynn-Williams DDEH: Environmental UV radiation: biological strategies for protection and avoidance. In *Astrobiology The quest for the conditions of life*. Edited by Gerda Horneck CB-K. Berlin: Springer; 2002:245–260.
57. Deamer DW: Boundary structures are formed by organic-components of the Murchison Carbonaceous Chondrite. *Nature* 1985, **317**(6040):792–794.
58. Cavalier-Smith T: The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* 2002, **52**(Pt 1):7–76.
59. Cavalier-Smith T: Cell evolution and Earth history: stasis and revolution. *Philos Trans R Soc Lond B Biol Sci* 2006, **361**(1470):969–1006.
60. Lombard J, Lopez-Garcia P, Moreira D: The early evolution of lipid membranes and the three domains of life. *Nat Rev Microbiol* 2012, **10**(7):507–515.
61. Myllykallio HSS, Grosjean H: Folate-Dependent Thymidylate-Forming Enzymes: Parallels between DNA and RNA Metabolic Enzymes and Evolutionary Implications. In *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Edited by Henri G. Austin (TX): Landes Bioscience; 2009:275–288.
62. Escartin F, Skouloubris S, Liebl U, Myllykallio H: Flavin-dependent thymidylate synthase X limits chromosomal DNA replication. *Proc Natl Acad Sci U S A* 2008, **105**(29):9948–9952.
63. Myllykallio H, Lipowski G, Leduc D, Filee J, Forterre P, Liebl U: An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* 2002, **297**(5578):105–107.
64. Chan CX, Bhattacharya D, Reyes-Prieto A: Endosymbiotic and horizontal gene transfer in microbial eukaryotes: Impacts on cell evolution and the tree of life. *Mob Genet Elements* 2012, **2**(2):101–105.
65. Nygaard P: Utilisation of preformed purine bases and nucleosides. In *Metabolism of Nucleotides, Nucleosides and Nucleobases in Microorganisms*. Edited by Munch-Peterson A. London: Academic Press, 1983; 27–93.
66. Bick MD, Davidson RL: Total substitution of bromodeoxyuridine for thymidine in the DNA of a bromodeoxyuridine-dependent cell line. *Proc Natl Acad Sci U S A* 1974, **71**(5):2082–2086.
67. Warren RA: Modified bases in bacteriophage DNAs. *Annu Rev Microbiol* 1980, **34**:137–158.
68. Marliere P, Patroux J, Doring V, Herdewijn P, Tricot S, Cruveiller S, Bouzon M, Mutzel R: Chemical evolution of a bacterium's genome. *Angew Chem Int Ed Engl* 2011, **50**(31):7109–7114.
69. Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, Tissier C, Zhang P, Karp PD: MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 2006, **34**(Database issue):D511–D516.
70. Claudel-Renard C, Chevalet C, Faraut T, Kahn D: Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* 2003, **31**(22):6633–6639.
71. Bairoch A: The ENZYME database in 2000. *Nucleic Acids Res* 2000, **28**(1):304–305.
72. Mavromatis K, Ivanova NN, Chen I-MA, Szeto E, Markowitz VM, Kyrpidis NC: The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009, **1**(1):63.
73. Wilson D, Madera M, Vogel C, Chothia C, Gough J: The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res* 2007, **35**(Database issue):D308–D313.
74. Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C: SCOP: a structural classification of proteins database. *Nucleic Acids Res* 2000, **28**(1):257–259.

doi:10.1186/1471-2164-15-800

Cite this article as: Armenta-Medina et al.: Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *BMC Genomics* 2014 **15**:800.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

