00361

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

DIVISION DE ESTUDIOS DE POSGRADO

"EVOLUCION MOLECULAR DE LAS RUTAS DE SALVAMENTO DE NUCLEOTIDOS DE PURINAS"

T E S I S

QUE PARA OBTENER EL GRADO ACADEMICO DE

MAESTRO EN CIENCIAS (BIOLOGIA)

P R E S E N T A

ARTURO CARLOS II BECERRA BRACHO

DIRECTOR DE TESIS:

DR. ANTONIO LAZCANO ARAUJO REYES

MEXICO, D. F. 1997

AGRADECIMIENTOS

De manera muy especial agradezco al Dr. Antonio Lazcano, la dirección de los trabajos aquí presentados, pero sobre todo su apoyo y amistad que me brindo durante el desarrollo de esta tesis.

De igual manera a los integrantes del laboratorio de microbiología, Sara, Amanda, Ana, Josetxu, Luis, Ulises, Ervin, Javier y Claudia por su ayuda y paciencia.

A los sinodales, por sus comentarios al trabajo.

A los integrantes del centro de informática (Enrique y Aquiles), por su asesoría con la red.

RESUMEN

Los caracteres moleculares como el 16/18s rRNA han demostrado el origen monofilético de todos los organismos y por ende, la existencia de un ancestro común a todos ellos (Woese 1983) sin embargo, es importante reconocer los límites de este enfoque tanto en la reconstrucción de filogenias y las características del cenancestro, como respecto a la poca o nula información que nos proporcionan sobre el origen de la vida mismo. No reconocerlo puede llevar a conclusiones prematuras, como ocurrió con la definición original del progenote sugerida por Woese y Fox (1977), o con la caracterización del ancestro como una célula con genoma de RNA y que supuestamente dependía heterotróficamente de nucleótidos de origen abiótico, tal como lo sugirieron Mushegian y Koonin (1996).

Por otro lado, en el estudio de la evolución de rutas metabólicas, la información que generan los marcadores moleculares puede ser significativo. Los nucleótidos de purinas se forman *de novo* por medio de una ruta metabólica de origen monofilético, pero también son sintetizados reutilizando bases y nucleósidos de purinas por medio de diferentes rutas de salvamento. El análisis filogenético de las enzimas que participan en las rutas de salvamento de nucleótidos de purinas generó evidencias sobre eventos de duplicación genica. Tres grupos monofiléticos de enzimas que catalizan reacciones mecánicamente similares fueron encontrados en este trabajo: (a) adenina-, xantina-, hipoxantina-, y guanina-fosforibosiltransferasa son homólogas entre si y además con las nucleósido fosforilasas; (b) el grupo formado por la adenina deaminasa, adenosina deaminasa, adenosina monofosfato deaminasa y adenilsuccinato sintasa; y (c) el par formado por la guanina reductasa y la inosina monofosfato deshidrogena. Estos resulatdos apoyan la idea de que las rutas ancestrales fueron mediadas por enzimas que poseían una baja especificidad al substrato (Waley, 1969; Ycas, 1974; Jensen, 1976), que pudieron participar en rutas metabólicas que actualmente no se encuentran directamente conectadas, como la biosíntesis de histidina y pirimidinas. Los nucleósidos y nucleótidos son compuestos poco estables por lo que es improbable su presencia en condiciones prebióticas. Por lo tanto las rutas de salvamento tuvieron que surgir después de que se desarrolló la biosíntesis de azúcares fosforiladas. En este trabajo de tesis también se presenta un esquema hipotético primitivo sobre la biosíntesis de nucleótidos, basado en la habilidad que tiene la adenina fosforibosiltransferasa de catalizar la condensación de 4-aminoimidazol-5-carboxamida (AICA) con PRPP.

INTRODUCCIÓN

1. Filogenia molecular y el ancestro universal

El análisis filogenético de las secuencias del 16/18S rRNA no solamente permitió a Woese y Fox (1977) caracterizar a las arqueobacterias como una rama monofilética claramente definida, sino que también mostró que todos los organismos conocidos forman parte de un árbol en el cual es posible distinguir otros dos grandes linajes, formados por las eubacterias y el nucleocitoplasma eucarionte. Este árbol evolutivo sin raíz, que resultó de la comparación de las secuencias de genes ortólogos de rRNA, se trifurca a partir de un ancestro común, al que Woese y Fox (1977) denominaron *progenote*. Debido a que no se ha descubierto un organismo que pueda servir como grupo externo a estos tres linajes celulares, el *progenote* fue definido no sólo como el ancestro común a las eubacterias, las arqueobacterias y los eucariontes, sino también como una entidad hipotética primitiva en la que la separación de fenotipo y genotipo aún no había tenido lugar (Woese y Fox, 1977). Años mas tarde, Woese (1983, 1987) continuó desarrollando su hipótesis y propuso que el progenote era un sistema donde el material hereditario estaba constituido por moléculas fragmentadas de RNA que aún no estaban integradas en un solo polímero genético (Woese, 1983, 1987).

No todos aceptaron la posibilidad de que el último ancestro común fuese, en efecto, un progenote. A partir del análisis de las secuencias de tRNAs de los tres linajes celulares, Fitch y Upper (1987) sugirieron que el ancestro común a estos ya poseía un código genético equivalente al de las células contemporáneas y propusieron que el árbol del RNA se trifurcaba no a partir de un *progenote* sino de un organismo complejo al que denominaron *cenancestro*. Por otra parte,

la comparación de las secuencias homólogas comunes a organismos de los tres linajes permitió proponer que el llamado *progenote* era, en realidad, una célula procarionte dotada de los mismos rasgos biológicos de una bacteria contemporánea (Lazcano *et al.*, 1992; Lazcano, 1995). A pesar de que no se puede excluir del todo ni la posibilidad de que hayan ocurrido fenómenos de transporte horizontal entre los tres linajes, ni de que hayan sucedido pérdidas secundarias en la divergencia de las ramas, la caracterización del ancestro común a estos tres grupos como una célula procarionte compleja sugiere la existencia de una fase de evolución biológica previa a su trifurcación .

Como es sabido, no fue sino hasta que se utilizaron conjuntos de genes parálogos que se habían duplicado antes de la separación de los tres linajes, cuando se pudo comenzar a construir árboles universales con raíz, la cual se ha ubicado en la rama eubacteriana (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989). Aunque la idea de que las eubacterias corresponden al fenotipo mas antiguo de todas las formas actuales de vida ha ido ganando una aceptación creciente (Brown y Dolittle, 1995), es igualmente cierto que existen anomalías aún no explicadas, entre las que se incluyen las filogenias construídas con secuencias de glutamato deshidrogenasas, glutamino sintetasas (Forterre *et al.*, 1993), cabamoil-fosfato sintetasas, proteínas de choque térmico y otras mas. Ello a llevado a sugerir que en el pasado pudo haber ocurrido un transporte masivo de genes entre los ancestros de bacterias gram positivas y las arqueobacterias (Gogarten, 1994).

Sin embargo, algo que pasó desapercibido para muchos, fue el hecho de que las duplicaciones parálogas ancestrales nos permiten asomarnos a épocas muy tempranas de la evolución anteriores al *cenancestro* mismo y describir, así sea en forma parcial, sistemas biológicos mas simples. Por ejemplo, la presencia

de dos conjuntos parálogos de factores de elongación (Iwabe *et al.*, 1989) y del caracter homólogo de las subunidades α y β de las ATPasas tipo F en los tres linajes (Gogarten *et al.*, 1989), permite reconocer una fase ancestral en la que la síntesis de proteínas requería de un solo factor de elongación, y en donde las ATPasas poseían habilidades reguladoras limitadas (Lazcano 1993, 1994). Este tipo de análisis, que se ha podido ir extendiendo a los conjuntos de genes parálogos que codifican a las DNA polimerasas, las aminoacil-tRNA sintetasas, las glutamino-amido transferasas y otras enzimas más, ha permitido reconocer una etapa temprana de la evolución biológica que no había sido caracterizada previamente. Así, aunque aún estamos lejos de comprender los eventos que tuvieron lugar entre el mundo del RNA y el cenancestro, el análisis de los genes parálogos comunes a los tres linajes permite no solamente detectar otros candidatos para enraizar árboles universales, sino también para describir, aunque sea de manera parcial, la evolución de la especificidad enzimática y la forma en que se fueron ensamblando las rutas metabólicas antes de la separación de eubacterias, arqueobacterias y eucariontes (Fani *et al.*, 1995).

2. El papel de las duplicaciones génicas en la evolución de rutas metabólicas

El primer intento por explicar el origen de las rutas metabólicas proviene del trabajo de Horowitz (1945), quien sugirió que el desarrollo y evolución de las vias biosintéticas es el resultado de una adquisición secuencial de enzimas, pero en un orden inverso al que actualmente poseen en una vía dada. Esta propuesta, basada en las ideas de Oparin (1938), establece una conexión evolutiva entre el ambiente primitivo y la emergencia del metabolismo biológico, al sugerir que los intermediarios bioquímicos de las rutas metabólicas básicas estaban ya presentes en el medio prebiótico. Tiempo después, el

descubrimiento de los operones llevó a Horowitz (1965) a desarrollar su idea ya conocida para entonces como la hipótesis retrógrada, al proponer que el agrupamiento de algunos genes biosintéticos, es el resultado precisamente de una sucesión temprana de duplicaciones génicas en tandem.

Sin embargo, es fácil encontrar una serie de objeciones a la hipótesis retrógrada. La mayoría de los intermediarios metabólicos son químicamente inestables y su acumulación en la sopa primitiva resulta difícil de explicar. Además muchos de estos intermediarios son compuestos fosforilados, los cuales no podrían permear fácilmente las membranas primordiales en ausencia de un sistema de transporte especializado. Por otro lado, la aplicación de la hipótesis retrograda implicaría que el DNA se originó antes que el RNA, pero existen muchas evidencias que sugieren lo contrario. Finalmente, en el análisis de los diferentes operones y enzimas de *Escherichia coli*, *Bacillus subtilis*, *Mycoplasma genitalium*, y *Haemophilus influenzae* y otros organismos no ha encontrado apoyo a favor de esta hipótesis (Islas *et al.*en prep.).

Una propuesta alternativa sobre el papel de la duplicación génica en el establecimiento de rutas anabólicas fue desarrollada independientemente por Waley (1969), Ycas (1974) y Jensen (1976). Según estos autores, las rutas biosintéticas fueron ensambladas por un mecanismo de "patchwork" o "bricollage" en que participaron enzimas primitivas con una baja especificidad al sustrato. De acuerdo con esta idea, las rutas metabólicas estaban catalizadas por enzimas que podían utilizar diferentes substratos parecidos. De esta manera, la duplicación génica y la divergencia evolutiva de las secuencias resultantes fue el mecanismo que incrementó el tamaño de los genes y aumentó la especificidad de las enzimas al substrato.

Si el mecanismo de ensamble propuesto por la hipótesis del *patchwork* funcionó en la evolución de rutas metabólicas, deben de existir huellas de dicho proceso en las secuencias nucleotídicas de las enzimas que participan en los pasos enzimáticos de los metabolismos. Por ejemplo, en la biosíntesis de histidina, el gen *hisG* que codifica para la ATP fosforibosiltranferasa, es homólogo tanto a los genes de otras fosforibosiltransferasas, como a los de nucleosidasas, y las proteínas codificadas por los genes *hisC* y *hisH* son homólogas a las demás aminotransferasas y glutaminamidotransferas tipo G, respectivamente (Fani, *et al*,1995). Ello demuestra que esta ruta anabólica se emsambló a partir del reclutamiento de genes que codificaban para enzimas menos específicas y pertenecientes a grandes familias.

¿Es el anterior un caso aislado? La homología entre la ATP fosforibosiltransferasa que participa en la biosíntesis de la histidina y las demás fosforibosiltransferasas (Bork *et al* 1995), incluyendo las que participan en el anabolismo de nucleótidos, sugiere que este mecanismo esta muy generalizado. Por ello, nos propusimos estudiar el metabolismo de purin-nucleótidos para analizar el papel que la duplicación génica ha jugado en su evolución temprana.

La síntesis *de novo* de ribonucleótidos purínicos es una ruta metabólica de amplia distribución filogenética (Henderson y Paterson, 1973). Ello sugiere que dicha vía estaba presente en el cenancestro. En contraste, las rutas de salvamento presentan una alta variedad tanto en su distribución biológica como en sus estrategias metabólicas (Becerra y Lazcano, 1997).

* Técnica textil, que emplea pedazos de diferentes telas, para la conformación de otra.

Estas rutas participan en la regulación intracelular de la concentración de nucleótidos, pero sobre todo en la reutilización de bases libres y nucleósidos de purina que son (mayoritariamente) productos de degradación metabólica que se perderían de no ser reciclados por este proceso (Berens *et al.*, 1995). Este reciclaje se puede llevar a cabo de varias maneras: (a) por medio de PRTasas, lo que genera una conversión directa de bases en ribonucleótidos; (b) bajo la reacción de quinasas, que promueven una conversión directa de nucleósidos a nucleótidos; (c) vía una nucleósido fosforilasa, que cataliza una conversión (reversible) de bases a nucleósidos; o (d) por medio de una desaminación de adenina y guanina lo que produce hipoxantina y xantina, respectivamente. Estos mecanismos no son excluyentes entre sí, y su presencia varia en los distintos linajes y aún entre especies cercanas.

Las rutas de salvamento son un proceso metabólico esencial en los seres vivos. En algunas especies de bacterias, la degradación fermentativa de purinas "salvadas" por este medio, así como de nucleósidos, constituyen una fuente importante de carbono, nitrógeno y energía (Munch-Peterson, 1983). Por otra parte, las rutas de salvamento constituyen una alternativa esencial para un grupo de parásitos obligados que han perdido en forma total o parcial la síntesis *de novo* de purinas, convirtiéndose en organismos estrictamente dependientes tanto de purinas exógenas como de nucleósidos y nucleótidos. Dicho grupo polifilético incluye a bacterias tales como *Mycoplasma genitalium*, eucariontes amitocondriales como *Giardia lamblia* y *Trichomonas vaginalis*, y otros protistas como cinetoplástidos, helmintos parásitos como *Schistosoma mansoni*, así como tejidos animales especializados como el cerebral (Tham *et al.*, 1993; Berens *et al.*, 1995; Tatusov *et al.*, 1996).

Debido a que muchos aceptan hoy en día que los primeros seres vivos se formaron a partir de compuestos orgánicos que se encontraban disponibles en el ambiente primitivo (Oparin, 1938), resulta tentador suponer que la naturaleza heterótrofa de estos primeros organismos hacía que su crecimiento y reproducción dependieran de nucleótidos y otros compuestos presentes en la sopa primitiva (Pennissi, 1996; Mushegian y Koonin, 1996). Sin embargo, los problemas asociados con la síntesis y la acumulación prebiótica de nucleótidos no sólo dificulta esta posibilidad, sino que limita la aplicación de la hipótesis retrógrada para explicar la biosíntesis de purinas (Keefe *et al.*, 1995). Ello sugiere que las rutas de salvamento surgieron después que la biosíntesis de azúcares fosforilados y nucleótidos de purinas.

Se ha sugerido que las PRTasas que actualmente forman parte de la rutas de salvamento, mediaban la unión de purinas con 5-fosfo-α-D-ribosil-1-pirofosfato (PRPP), para formar los nucleótidos-5′ correspondientes (Zubay, 1993). En uno de los trabajos de esta tesis, se sugiere un esquema adicional, basado en un mecanismo semi-enzimático, donde la PRTasa cataliza la reacción entre el compuesto prebiótico 5-amino-4-imidazolcarboxamida (AICA) y la PRPP, produciendo el ribótido 5-amino-4-imidazolcarboxamida (AICAR), que es un intermediario en la síntesis *de novo* de purin-nucleótidos. Debido a que esta ruta hipotética requiere de la PRPP, para la cual no hay una síntesis prebiótica, es probable que la vía propuesta surgió luego del anabolismo de azúcares fosforiladas. El modelo propuesto tiene la ventaja adicional de no requerir de los ocho primeros pasos de la síntesis *de novo*.

Recientemente el análisis y comparación de secuencias de proteínas ha demostrado que buena parte de las enzimas de salvamento pertenecen a familias involucradas principalmente en metabolismo de nucleótidos (Mushegian y

Koonin, 1994; Bork *et al.*, 1995; Holm y Sander, 1996). Sin embargo, a la fecha está por analizarse su relación evolutiva en términos de origen y desarrollo de las rutas metabólicas, y su posible participación en condiciones primitivas. Por otra parte, aunque el conocimiento sobre la distribución biológica y la diversidad de las rutas de salvamento es aún incompleto, en este trabajo se ha podido demostrar que las rutas de salvamento de nucleótidos de purinas surgieron de un proceso de *patchwork*, en donde las duplicaciones parálogas jugaron un papel importante en el incremento del tamaño y la complejidad de los genomas ancestrales. Como se discute en los textos anexos, los resultados de este trabajo no sólo nos permiten generalizar las conclusiones establecidas previamente para el caso de biosíntesis de la histidina (Fani *et al.*, 1995), sino que permiten comenzar a establecer una conexión entre la química prebiótica y el origen de ciertas rutas metabólicas mas allá de las premisas pioneras de Horowitz (1945).

Estructura de la tesis.

Este trabajo de tesis está conformado por tres manuscritos, de los cuales dos son artículos y el tercero es un capítulo de libro en prensa. Todos ellos han sido aceptados en publicaciones internacionales arbitradas. Los textos son:

1.- Becerra-Bracho A., Silva E., Velasco A.M., y Lazcano A:
Molecular biology and the reconstruction of microbial phylogenies: des liaisions dangereuses?. In ***Physics of the origin of life*** editado por V. Stefan y publicado por Interdisciplinary Physics Series of la Jolla International School of Physics, Institute of Advanced Physics, American Institute of PhysicsPress, La Jolla CA. (en prensa).

2.- Becerra A., Islas S., Leguina, J., Silva E. y Lazcano A.
Polyphyletic gene losses can bias backtrack characterizations of the cenancestor in ***Journal Molecular Evolution*** (en prensa).

3.- Becerra A. y Lazcano A:
The role of gene duplication in the evolution of purine nucleotide salvage pathways, enviado a la revista ***Origin of life and the Evolution of the Biosphere*** (enviado).

MOLECULAR BIOLOGY AND THE RECONSTRUCTION OF MICROBIAL PHYLOGENIES: DES LIAISONS DANGEREUSES?

A. Becerra-Bracho, E. Silva, A. M. Velasco, and A. Lazcano*

Departamento de Biología
Facultad de Ciencias, UNAM
Apdo. Postal 70-407
Cd. Universitaria, México 04510, D.F.
MEXICO

FAX No. 525/616-0451 or 525/622-4828

* corresponding author

1. INTRODUCTION

Only forty years after the DNA double helix model was first suggested, molecular biology has become one of the most provocative, rapidly developing fields of of scientific research, that has led not only to tantalizing new findings on processes and mechanisms at the molecular level, but also to major conceptual revolutions in life sciences. Is there any hope of developing methodological approaches and theoretical frameworks not only to make sense of the overwhelming growing body of data that this relatively new field is producing, but also to use them to develop a more integrative, truly multidisciplinary understanding of biological phenomena? As Peter Bowler wrote a few years ago, Charles Darwin and his followers were accutely aware that "evolutionism's strength as a theory came fom its ability to make sense out of a vast range of otherwise meaningless facts" (Bowler, 1990). This situation has not changed. Attempts to explain the basic properties of th eliving beings as the result of complex systems in whose emergence self-organization processes have played a major role are now fashionable (Kauffman, 1993). However, the unifying powers of evolutionary biology have not been diminished. In fact, they probably represent one of the most promising possibilities of overcoming the perils of reductionism that have plagued molecular biology since its inception.

Molecular approaches to evolutionary issues are almost a century old. The possibility of developing a sucessful blending between them may have been first suggested by the American-born British biologist and physician George H. F. Nuttall, who in 1904 published a book summarizing the results of the detailed comparisons of blood proteins that he had used to reconstruct the evolutionary relationships of animals. "In the absence of palaeontological evidence", wrote Nuttall (1904), "the question of the interrelation-ship amongst animals is based upon similarities of structure in existing forms. In judging of theses similarities, the subjective element may largely enter, in evidence of which we need but look at the history of the classification of the Primates." Such subjective element, Nuttall believed, could be successfully overcomed by constructing a phylogeny based not on form but on the inmunological reactions of blood-related proteins.

Although the comparative analysis of biochemical properties, metabolic pathways and, in few cases, morphological characteristics, had provided some useful insights on the evolutionary relationships among certain microorganisms, until a few years ago the reconstruction of bacterial phylogenies and the understanding of microbial taxonomy were both viewed with considerable skepticism. This situation has undergone dramatic changes with the recognition that proteins and nucleic acid sequences are historical documents of unsurpassed evolutionary significance (Zuckerkandl and Pauling, 1965), and has led to a radical renovation of the phylogeny, classification, and systematics of prokaryotic and eukaryotic microbes (Woese, 1987).

But these changes have also sparked new debates, and have led to an increased appreciation that the scope and limits of molecular cladistic methodologies require clarification. As shown by the current controversies on the characteristics of the first organisms, the origin of the different components of the eukaryotic cell, the tempo and mode of microbial evolution, and the soundness of traditional taxonomic systems, the development of the full potential of molecular cladistics will depend not only on methodological refinements to improve the algorithms used for reconstructing evolutionary history from molecular data, but also the critical reexamination of its theoretical framework, which includes a number of central concepts, most of which were grafted from classical evolutionary theory into molecular biology. The purpose of this chapter is to discuss some of these issues, and to review briefly some of the major contributions that they have promoted in our understanding of previously uncharacterized early periods of biological evolution.

2. A EUKARYOTE IS A EUKARYOTE IS A EUKARYOTE?

The awareness that genomes are extraordinarily rich historical documents from which a wealth of evolutionary information can be retrieved has widened the range of phylogenetic studies to previously unsuspected heights. The development of rapid nucleic acid sequencing techniques, combined with

the simultaneous and independent blossoming of computer science, has led not only to an explosive growth of sequence databases and new sophisticated tools for their exploitation, but also to the recognition that different macromolecules may be uniquely suited as a molecular chronometers in the construction of nearly universal phylogenies.

A major achievement of this approach has been the evolutionary comparison of small subunit ribosomal RNA (rRNA) sequences, which has allowed the construction of a trifurcated, unrooted tree in which all known organisms can be grouped in one of three major (apparently) monophyletic cell lineages: the eubacteria, the archaebacteria, and the eukaryotic nucleocytoplasm, now referred to as new taxonomic categories, i.e., the domains *Bacteria*, *Archaea*, and *Eucarya*, respectively (Woese et al., 1990). There is strong evidence that the identification of these lineages is not an artifact based solely upon the reductionist extrapolation of information derived from one single molecule. It is true that the congruence between rRNA genes and other molecules is not always ideal, and anomalous phylogenies have been reported (Rivera and Lake, 1992; Tiboni et al., 1993; Gupta and Golding, 1993). At the time being there is no general explanation to account for these preculiar topologies, and the possibility that we may have to restrict ourselves to empirical characterizations of such cases should be kept in mind. However, a large variety of phylogenetic trees constructed from DNA and RNA polymerases, elongation factors, F-type ATPase subunits, heat-shock and ribosomal proteins, and an increasingly large set of genes encoding enzymes involved in biosynthetic pathways, have confirmed the existence of the three primary cellular lines of evolutionary descent (Doolittle and Brown, 1994).

The ensuing tripartite taxonomic description of the living world fostered by C. R. Woese and his followers has been disputed by a number of workers, who contend that both eubacteria and archaebacteria are *bona fide* prokaryotes, regardless of the pecularities that separate that separate them at the molecular level (Mayr, 1990; Margulis and Guerrero, 1991; Cavalier-Smith, 1992). Furthermore, because of their very nature, molecular dichotomous phylogenetic trees cannot be drawn which include anastomozing branches corresponding to the lineages which gave rise to the different components of eukayotic cells. Accordingly, Margulis and Guerrero

(1991) have argued that although molecular cladistics is now a prime force in systematics, phylogenetically accurate taxonomic classifications should be based not only on the evolutionary comparison of macromolecules, but also on metabolic pathways, chromosomal cytology, ultrastructural morphology, biochemical data, life cycles, and, when available, paleontological and geochemical evidence.

By showing the evolutionary proximity of mitochondria and chloroplasts to purple bacteria and cyanobacteria, respectively, molecular phylogenies have confirmed the endosymbiotic theory. Moreover, a number of trees also suggest that a major portion of the eukaryotic nucleocytoplasm originated from an archaebacteria-like cell whose descendants form the monophyletic eucaryal branch (Gogarten-Boekels and Gogarten, 1994). As asserted by Woese and his collaborators, although the presence of endosymbionts is of critical importance to the eukaryotes, it is undeniable that the latter "have a unique, meaningful phylogeny" (Wheelis et al., 1992). While such view assumes an absolute continuity between the nucleocytoplasm and its direct ancestor, the holistic arguments advocated by Margulis and Guerrero (1991), Cavalier-Smith (1992), and others, emphasize the evolutionary emergence of an novel type of cell as a result of endosymbiotic events. According to the latter, the key transitional event leading to nucleated cells was the evolutionary acquisition of heritable intracellular symbionts, and the eucaryal branch does not represent eukaryotic cells as a whole, any more than fungal hyphae or phycobionts like the *Trebouxia* algal cells exhibit, by themselves, all the phenotypic and genetic characteristics of a lichen thallus.

Of course, antagonistic taxonomies have coexisted more or less peacefully along the history of biology. However, the urgent need to critically revise current classificatory systems cannot be underscored. Modern taxonomic schemes need to acknowledge not only the existence of three major cell lineages, but also the eukaryotic divergence patterns, which appear to be the result of rapid bursts of speciation (Sogin, 1994). Any such modifications in biological classification require the recognition of the functional and anatomical continuity between the eukaryotic cytoplasm and the intranuclear environment, as well as the likelihood that the evolution of membrane-bounded nuclei is indeed a byproduct of permanent intracellular associations.

In fact, even amitochondrial eukaryotes such as *Giardia* and *Trychomonas* appear to have permanent intracellular bacterial endosymbionts (Nemanic et al., 1979; Feely et al., 1988). These pre-mitochondrial cells, which may include the microaerophilic, amitotic, multinucleated giant amoeba *Pelomyxa palustris*, are all located in the oldest branches of the eucarya, and contain several types of intracellular prokaryotes which may be the functional equivalents of mitochondria. The ubiquity of endosymbionts suggests that they may have played a critical rǫle in the evolutionary development of nucleated cells. This hypothesis is amenable to observational and experimental designs, and may be supported by studying the possible bacterial affinities of membrane-bounded hydrogenosomes that are known to multiply by binary division in the *Trychomonas* cytoplasm (Müller, 1988), as well as by searching for prokaryotic endosymbionts in species of Parabasalia, Retortomonads, Diplomonads, Calonymphids, and other protist taxa which apparently evolved prior to mitochondrial acquisition.

3. THE ROOT OF THE TREE OR THE TIP OF THE TRUNK?

The construction of the unrooted rRNA tree showed that no single major branch predates the other two, and all three derive from a common ancestor. It was thus concluded that the latter was a progenote, which was defined as a hypothetical entity in which phenotype and genotype still had an imprecise, rudimentary linkage relationship (Woese and Fox, 1977). According to this view, the differences found among the transcriptional and translational machineries of eubacteria, archaebacteria, and eukaryotes, were the result of evolutionary refinements that took place separately in each of these primary banches of descent after they have diverged from their universal ancestor (Woese, 1987).

From an evolutionary point of view it is reasonable to assume that at some point in time the ancestors of all forms of life must have been less complex than even the simpler extant cells, but our current knowledge of the characteristics shared between the three lines has shown that the conclusion that the last common ancestor was a progenote was premature. This interpretation, based on rRNA-based trees for which no outgroups have been

discovered, has been definitively superseded (Woese, 1993). A partial description of the last common ancestor of eubacteria, archaebacteria, and eukaryotes may be inferred from the distribution of homologous traits among its descendants. The set of such genes that have been sequenced and compared is still small, but the sketchy picture that has already emerged suggests that the most recent common ancestor of all extant organisms, or *cenancestor*, as defined by Fitch and Upper (1987), was a rather sophisticated cell with at least (a) DNA polymerases endowed with proof-reading activity; (b) ribosome-mediated translation apparatus with an oligomeric RNA polymerase; (c) membrane-associated ATP production; (d) signalling molecules such as cAMP and insulin-like peptides; (e) RNA processing enzymes; and (f) biosynthetic pathways leading to amino acids, purines, pyrimidines, coenzymes, and other key molecules in metabolism (Lazcano et al., 1992; Benner et al., 1993).

These traits are far to numerous and complex to assume that they evolved independently or that they are the result of massive multidirectional horizontal transfer events which took place before the earliest speciation events recorded in each of the three lineages. Their presence suggests that the cenancestor was not a direct, immediate descendant of the RNA world, a protocell or any other pre-life progenitor system. Very likely, it was already a complex organism, much akin to extant bacteria, and must be considered the last of a long line of simpler earlier cells for which no modern equivalent is known.

Unfortunately, the characteristics of evolutionary predecessors of the cenancestor cannot be inferred from the plesiomorphic traits found in the space defined by rRNA sequences. Although trees constructed from such universally shared characters appear to be free of internal inconsistencies, the lack of outgroups leads to topologies that specify branching relationships but not the position of the ancestral phenotype. Thus, such trees cannot be rooted. This phylogenetic *cul-de-sac* may be overcomed by using paralogous genes, which are sequences that diverge not through speciation but after a duplication event. As noted almost twenty years ago by Schwartz and Dayhoff (1978), rooted trees can be constructed by using one set of paralogous genes as

an outgroup for the other set, a rate-independent cladistic methodology that expands the monophyletic grouping of the sequences under comparison.

This approach was used independently a few years ago by Iwabe et al (1989) and Gogarten et al (1989), who analyzed paralogous genes encoding (a) the two elongation factors (EF-G and EF-Tu) that assist in protein biosynthesis; and (b) the alpha and beta hydrophilic subunits of F-type ATP synthetases. Using different tree-constructing algorithms, both teams independently placed the root of the universal trees between the eubacteria, on the one side, and archaebacteria and eukaryotes on the other. Their results imply that eubacteria are the oldest recognizable cellular phenotype, and suggests that specific phylogenetic affinities exist between the archaea and the eucarya.

This branching order, which was promptly adopted by Woese et al (1990), appears to be consistent with structural and functional similarities which are known to exist in the translation and replication machineries of both archaebacteria and eukaryotes (Kletzin, 1992; Ouzonis and Sander, 1992; Slesarev et al., 1993; Kaine et al., 1994). However, the issue is far from solved. No support for a particular topology was detected when mean interdomain distance analysis was used to analize a set of approximately forty genes common to the three lineages (Doolittle and Brown, 1994). The situation is further aggravated by the fact that the phylogenetic analysis of sets of ancestral paralogous genes other than the elongation factors and the ATPase hydrophilic subunits has challenged the conclusion that universal trees are rooted in the eubacterial branch (Benahenchou-Lahfa et al., 1993; Gupta and Singh, 1992; Forterre et al., 1993). While the products of genes involved in the transcription/translational molecular machinery of eukaryotes appear to be closer to those of archaebacteria than to eubacteria, other sequences such as those encoding heat-shock proteins and several enzymes suggest the existence of phylogenetic affinities between archaea and Gram positive bacteria. These observations have led to the suggestion that the pre-eukaryotic host was the outcome of a fussion event between eubacteria and archaebacteria, that took place prior to the endosymbiotic acquisition of mitochodnria and chloroplasts (Gupta, 1995).

The lack of congruency between different universal phylogenies may be the result not only of the statistical problems involved in the aligment and comparison of a large number of sequences that may have diverged more than 3.5×10^9 years ago, but also of even older additional paralogous duplications (Forterre et al., 1993), and of horizontal gene tranfer events (Hilario and Gogarten, 1993), both of which may be obscuring the natural relationships between the lineages. Given the likelihood that microbial phylogenetic analysis will increase its reliance on paralogous duplicates to define outgroups and character polarities (Sidow and Bowman, 1991), detailed studies should be devoted to assess the validity and limits of this cladistic methodology.

Minor differences in the basic molecular processes of the three main cell lines can be distinguished, but all known organisms, including the oldest ones, share the same essential features of genome replication, gene expression, basic anabolic reactions, and membrane-associated ATPase mediated energy production. The molecular details of these universal processes not only provide direct evidence of the monophyletic origin of all extant forms of life, but also imply that that sets of genes encoding the components of these complex traits were frozen a long time ago, i. e., major changes in them are very strongly selected against and are lethal. Biological evolution prior to the divergence of the three domains was not a continuous, unbroken chain of progressive transformation steadily proceeding towards the cenancestor. However, no evolutionary intermediate stages or ancient simplified version of the basic biological processes have been discovered in extant organisms.

Nevertheless, clues to the genetic organization and biochemical complexity of the earlier entities from which the cenancestor evolved may be derived from the analysis of paralogous sequences. Their presence in the three cell lineages implies not only that their last common ancestor was a complex cell already endowed, among others, with pairs of homologous genes encoding two elongation factors, two ATPase hydrophilic subunits, two sets of glutamate dehydrogenases, and the A and B DNA polymerases, but also that the cenancestor itself must have been preceded by simpler cells in which only one copy of each of these genes existed. In other words, Archean paralogous genes provide evidence of the existence of ancient organisms in which ATPases

lacked the regulatory properties of its alpha subunit, protein synthesis took place with only one elongation factor, and the enzymatic machinery involved in the replication and repair of DNA genomes had only one polymerase ancestral to both DNA polymerase I and II.

By definition, the node located at the bottom of the cladogram is the root of a phylogenetic tree, and corresponds to the common ancestor of the group under study. But names may be misleading. The recognition that basic biological processes like DNA replication, protein biosynthesis, and ATP production require today the products of pairs of genes which arose by paralogous duplications during the early Archean, implies that what we have been calling the root of universal trees is in fact the tip of a trunk of unknown length in which the history of a long (but not necessarily slow) series of archaic evolutionary events may still be recorded. The inventory of paralogous genes that duplicated during this previously unchacterized stage of biological evolution appears to include, in addition to elongation factors, ATPase subunits, and DNA polymerases, the sequences encoding heat shock proteins, ferredoxins, dehydrogenases, DNA topoisomerases, several pairs of aminoacyl-tRNA synthetases, and enzymes involved in nitrogen metabolism and amino acid biosynthesis. It is noteworthy that this list includes also aspartate transcarbamoyl transferase, an enzyme which together with carbamyl phosphate synthetase (whose large subunit is itself the product of an internal, i.e., partial, paralogous duplication) catalyzes the initial steps of pyrimidine biosynthesis (García-Meza et al, 1995).

Thus, prior to the early duplication events that led to what may be a rather large number of cenancestral paralogous sequences, simpler living systems existed that lacked the large set of enzymes and the sophisticated regulatory abilities of contemporary cells. Although lateral transfer of coding sequences may be almost as old as life itself, gene duplication followed by divergence probably played a prominent role in the accretion of complex genomes, and may have led to a rapid rate of microbial evolution. If it is assumed that the rate of gene duplicative expansion of ancient cells was comparable to today's present values, which are of 10^{-5} to 10^{-3} gene duplications per gene per cell generation (Stark and Wahl, 1984), the maximum time required to go from an hypothetical 100 gene organism to one endowed with a filamentous

cyanobacterial-like genome of approximately 7000 genes would be less than ten million years (Lazcano and Miller, 1994).

Although the are no published data on the rate of formation of new enzymatic activities resulting from gene duplication events under either neutral or positive selection conditions, the role of duplicates in the generation of evolutionary novelties is well stablished. Once a gene duplicates, one of the copies may be free to accumulate non-lethal mutations and acquire new additional properties, which could lead into its specialization or recruitment into new role. Data summarized here supports the idea that primitive biosynthetic pathways were mediated by small, inefficient enzymes of broad substrate specificity (Jensen, 1976). Larger substrate ranges may had not been a disadvantage, since relatively unspecific enzymes may have helped ancestral cells with reduced genomes overcome their limited coding abilities (Ycas, 1974).

The discovery that homologous enzymes catalyzing similar biochemical reactions are part of different anabolic pathways supports the idea that enzyme recruitment took place during the early development of several basic anabolic pathways (Riley, Fani, Forterre, Labedan, and Lazcano, in prep.). Evolutionary tinkering of the products of duplication events apparently had a major role in metabolic evolution. This is supported by the analysis of the *Escherichia coli* sequence databases, that has shown that approximately 36% of the proteins whose sequence is available are the result of duplication events (Riley, 1993). Such high levels of redundancy represent an illuminating possibility and suggest that the wealth of phylogenetic information older than the cenancestor may be larger than realized, and its analysis may provide fresh insights into a crucial but largely undefined stage of early biological evolution during which major biosynthetic pathways emerged and became fixed.

There is a major exception to the above conclusion. True fungi, euglenids, and chytridiomycetes synthesize lysine via an eight-step pathway in which α-aminoadipate (AAA) is an intermediate. This route is different from the seven-step diaminopimelate pathway used by bacteria, plants, and most protists (Bhattacharjee, 1985). The phylogenetic distribution of these pathways

suggest that the diaminopimelate route is the most recent one. Accordingly, if the patchwork assembly of metabolic pathways (Jensen, 1976) is valid, then it can be predicted that the enzymes catalyzing the AAA-route should be homologous to those participating in other major biosynthetic pathways.

The recognition that enzyme recruitment may have played a major role in metabolic evolution leads, however, to assume some caution in phylogenetic inferences. Although in some cases metabolic pathways may be sucessfully used to assess the phylogenetic relationship of prokaryotes (DeLey, 1968; Margulis, 1993), the possibility that some of the enzymes of archaic pathways may have survived in unusual organisms (Keefe et al., 1994), or that important portions of extant metabolic routes may have been assembled by a patchwork process (Jensen, 1976), suggest that considerable prudence should be exerted when attempting to describe the physiology of truly primordial organisms by simple direct back extrapolation of extant metabolism.

4. MOLECULAR CLADISTICS AND THE ORIGIN OF LIFE: THE WEAK CONNECTION

"All the organic beings which have ever lived on this Earth", wrote Charles Darwin in the *Origin of Species*, "may be descended from some primordial form". Although the placement of the root of universal trees is a matter of debate, the development of molecular cladistics has shown that despite their overwhelming diversity and tremendous differences, all organisms are ultimately related and descend from Darwin's primordial ancestor. But what was the nature of this progenitor?

The heterotrophic hypothesis suggested by Oparin (1938) not only gave birth to a whole new field devoted to the study of the origin of life, but played a central role in shaping several influential taxonomic schemes and different bacterial phylogenies (Broda, 1975; Almassy and Dickerson, 1978; Schwartz and Dayhoff, 1978; Margulis 1993). Although the central role of glycolysis and the wide phylogenetic distribution of at least some of its molecular components are strong indications of its antiquity (Fothergill-Gilmore and Michels, 1993), it is no longer possible to support the *ad hoc* identification of

putative primordial traits to assume that the first living system was a *Clostridium*-like anaerobic fermenter (Schwartz and Dayhoff, 1978; Almassy and Dickerson, 1978) or a *Mycoplasma* type of cell (Wallace and Morowitz, 1973; Razin, 1978). Like vegetation in a mangrove, the roots of universal phylogenetic trees are sumerged in the muddy waters of the prebiotic broth, but how the transition from the non-living to the living took place is still unknown.

Indeed, we are still very far from understanding the origin and attributes of the first living beings, which may have lacked even the most familiar features in extant cells. For instance, protein synthesis is such an essential characteristic of cells, that it is frequently argued that its origin should be considered synonymous with the emergence of life itself. However, the discovery of the catalytic activities of RNA molecules has led considerable support to the possibility that during early stages of biological evolution living systems were endowed with a primitive replicating and catalytic apparatus devoid of both DNA and proteins (Lazcano, 1994a). The scheme may be even more complex, since RNA itself may have been preceded by simpler genetic macromolecules lacking not only the familiar 3',5' phosphodiester backbones of nucleic acids (Joyce et al., 1987), but perhaps even today's bases (Kolb et al., 1994).

Although the discovery that L-arginine can bind in a highly selective way to catalytic introns supports the hypothesis that the genetic code emerged in an RNA world (Yarus, 1993), the ultimate origin of the codon assigments remains an unsolved problem. Molecular cladistic analysis of the components of protein synthetic machinery (ribosomal proteins, elongation factors, aminoacyl tRNA synthetases, initiation and termination factors, etc.) may provide some clues on the late evolution of the genetic apparatus. Systematic studies of complete sets of aminoacyl-tRNA synthetases from the three cell linages are still required, but the comparison of the available sequence and tertiary structures (Cusack et al., 1990; Nagel and Doolittle, 1991) have confirmed that these keys enzymes fall into two functionally related but distinct classes that appear to have evolved independently.

Group I includes the valyl-, isoleucyl-, and the leucyl-tRNA synthetases. The topology of the available unrooted tree of isoleucyl-tRNA synthetase sequences from the three cell linaeages suggests a sisterhood relationship between the two prokaryotic domains (Doolittle and Brown, 1994) that is not consistent with the rooted trees calculated by Gogarten et al (1989) and Iwabe et al (1989). Nonetheless, the close phylogenetic relationship between the *Escherichia coli* isoleucyl-tRNA synthetase sequence, and the yeast valyl-tRNA synthetase (Jordana et al., 1987) suggests only that valine and isoleucine synthetases are the products of still another archaic set paralogous genes diverged before the cenanestor's time. Equally significant, this similarity also suggests that the ancestral synthetase may have been a non-specific enzyme unable to discriminate between valine and isoleucine.

Since valine and isoleucine are two sterically related amino acids, have similar codons, and share the same enzymes in the last four parallel reactions of their biosynthetic pathways, the phylogenetic proximity of their tRNA synthetases suggests that the concerted evolution of an entire set of genes took place, eventually leading to a less-ambiguous code. This interpretation, which is consistent with the coevolution theory of the genetic code suggested by Wong (1975), clearly fits into previous schemes suggesting that the genetic apparatus evolved from primitive inaccurate versions in which groups of chemically similar amino acids may have been activated by unspecific synthetases (Woese, 1965; Fitch and Upper, 1987). It is likely that the hydrolytic sites that provide these enzymes with an additional sieve that allows them to discriminate between valine and isoleucine (Norris and Berg, 1964), emerged after the evolution of valyl- and isoleucyl-tRNA synthetase specificities as a result of duplication events.

Although molecular cladistics may provide clues to some late steps in the development of the genetic code, it is difficult to see how the applicability of this approach can be extended beyond a threshold that corresponds to a period of cellular evolution in which protein biosynthesis was already in operation. Older stages are not yet amenable to molecular phylogenetic analysis. Although there have been considerable advances in the understanding of chemical processes that may have taken place before the emergence of the first living systems, life's beginnings are still shrouded in mystery. A cladistic

approach to this problem is not feasible, since all possible intermediates that may have once existed have long since vanished. The temptation to do otherwise is best resisted. Given the huge gap existing in current descriptions of the evolutionary transition between the prebiotic synthesis of biochemical compounds and the cenancestor (Lazcano, 1994b), it is naive to attempt to describe the origin of life and the nature of the first living systems from the available rooted phylogenetic trees.

Nevertheless, there have been several recent attempts to use macromolecular data to support claims on the hyperthermophily of the first living organisms and the idea of a hot origin of life. The examination of the prokaryotic branches of unrooted rRNA trees had already suggested that the ancestors of both eubacteria and archaebacteria were extreme thermophiles, i.e., organisms that grow optimally at temperatures in the range 90° C and above (Achenbach-Richter et al., 1987). Rooted universal phylogenies appear to confirm this possibility, since heat-loving bacteria occupy short branches in the basal portion of molecular cladograms (Stetter, 1994).

Such correlation between hyperthermophily and primitiveness has led support to the idea that heat-loving lifestyles are relics from early Archean high-temperature regimes that may have resulted from a severe impact regime (Sleep et al., 1989). It has also been interpreted as evidence of a high temperature origin of life, which according to these hypotheses took place in extreme environments such as those found today in deep-sea vents (Holm, 1992), or in other sites in which mineral surfaces may have fueled the appearance of primordial chemoautolithotrophic biological systems (Wächtershäuser, 1990).

Such ideas are not totally without precedent. The possibility that the first heterotrophs may have evolved in a sizzling-hot environment is in fact an old suggestion (Harvey, 1924). Despite their long genealogy, these hypotheses have not been able to bypass the problem of the chemical decomposition faced by amino acids, RNA, and other thermolabile molecules which have very short lifetimes under such extreme conditions (Miller and Bada, 1988). Although no mesophilic organisms older than heat-loving bacteria have been discovered, it is possible that hyperthermophily is a secondary

adaptation that evolved in early geological times (Sleep et al., 1989; Confalonieri et al., 1993; Lazcano, 1993). In fact, hyperthermophiles not only share the same basic features of the molecular machinery of all other forms of life; they also require a number of specific biochemical adaptations. Any theory on the hot origin of life must address the question of how such traits, or their evolutionary precedessors, arose spontaneously in the prebiotic environment. Such adaptations may include histone-like proteins, RNA modificating enzymes, and reverse gyrase, a peculiar ATP-dependent enzyme that twists DNA into a positive supercoiled conformation (Segerer et al., 1993; Confalonieri et al., 1993). Clues to the origin of hyperthermophily may be hidden in this list, and its evolutionary analysis may contribute to the understanding of the rather surprising phylogenetic distribution of the immediate mesophilic descendants of heat-loving prokaryotes, which shows that at least five independent abandonments events of hyperthermophilic traits took place in widely separated branches of universal trees, one of which corresponds to the eukaryotic nucleocytoplasm (García-Meza et al., 1995).

The antiquity of hyperthermophiles appears to be well stablished, but there is no evidence that they have a primitive molecular genetic apparatus. Thus, the most basic questions pertaining to the origin of life relate to much simpler replicating entities predating by a long series of evolutionary events the oldest recognizable heat-loving bacteria. Why hyperthermophiles are located at the base of universal trees is still an open question, but the possibility that adaptation to extreme environments is part of the evolutionary innovations that appeared in trunk of the tree cannot be entirely dismissed. The phylogenetic distribution of heat-loving bacteria is no evidence by itself of a hot origin of life, any more than the presence in the hyperthermophile archaeon *Sulfolobus solfataricus* of a gene encoding a thermostable B-type DNA polymerase endowed with 3'-5' exonuclease activity (Pisani et al., 1992) can be interpreted to imply that the first living organism had a DNA genome.

5. THE TIMESCALES OF MICROBIAL EVOLUTION

Because of their short generation times and to their tendency to engage in taxonomically promiscuous lateral transfer of genes, prokaryotes can adapt

rapidly to new environmental conditions and evolve at an extremely quick pace. New traits can become fixed in bacterial populations in timescales that are mere instants from a geological perspective. However, due perhaps to the extended intellectual prejudice that all evolutionary changes must proceed a in a step-wise, slow pace, it has been generally been assumed that the origin and early evolution of life were extremely slow processes requiring several billions of years (Oparin, 1938; Wald, 1954; Simpson, 1964; Cloud, 1968).

This view is no longer tenable. Late accretion impacts may have killed off life on our planet as late as 3.8×10^9 years ago (Sleep et al., 1989), but as shown by the Northwestern Australian fossil assemblages, an abundant, complex and highly diversified microbiota which may have included stromatolite-building cyanobacteria, was fluorishing only 300 million years after the period of intense bombardment had ended (Schopf, 1993). Since rRNA-based phylogenetic trees suggest that cyanobacteria are a late eubacterial clade (Woese, 1987), Archean life must have rapidly achieved levels of genetic organization, biochemical complexity, ecological diversity, and evolutionary potential, comparable to those of extant bacterial populations.

Although it is not possible to assign a precise chronology to the earliest branching events, there are several indications that the three domains of life could have already been distinct by the early Precambrian. The fact that the inside of the cell nucleus is a poorly oxygenated milieu in which no O_2-dependent metabolic pathways appear to exist (Joenje, 1989), may be interpreted as the retention of a primitive feature, thereby supporting the idea that the ancestor of the eukaryotic nucleocytoplasm branched off early in the history of planet. Of course, no mitochondrial-bearing eukaryotes could have survived in the oxygen-poor primitive environment, but the free-living ancestors of parasitic amitochondrial anaerobes like *Giardia* could have done so (Knoll, 1992).

While it is true that for many groups of plants and animals there is a remarkably well-established correlation between the number of amino acid differences and the time since their evolutionary divergence from a common ancestor, it is notoriously difficult to put microbial molecular evolution within a temporal framework. Base-pair substitution rates estimated from

paleontological information have been used to calibrate molecular trees, but the prokaryotic fossil record is so scanty that estimates on the chronological lengths of bacterial branches and the assigment of reliable dates to phylogenetic trees are prccluded.

Due to the inadequeacy of the bacterial fossil record, Ochman and Wilson (1987) attempted to obtain multiple calibration points for the eubacterial 16S rRNA lineage by linking the divergence of some of their branches with the geological record. The antiquity of the nitrogen-fixing *Rhizobium* and luminous *Photobacterium* bacteria, which have symbiotic associations with legumes and marine fishes, respectively, was estimated from the plant and vertebrate fossil record, for example. Approximately 1% base substitution per 50 million years was calculated (Ochman and Wilson, 1987). The living descendants of aphids have maternally inherited endosymbiotic proteobacteria that have radiated synchronously with their host, allowing a calculation of the temporal scale of eubacterial evolution (Moran et al., 1993). By calibrating the endosymbiotic bacterial tree against the fossil aphid record, Moran and her collaborators estimate constant rates of base substitution of 1 to 2% per 50 million years.

The above calculations suggest that the rates of eubacterial 16S rRNA substitution are constant (Ochman and Wilson, 1987; Moran et al., 1993), but other observations indicate that DNA substitution rates change at varying porportions in different organisms and in different times. Thus, calibration of the bacterial clock can be used to calculate the time of divergence of recent prokaryotes like the enterobacteria, for instance, but to extrapolate it back into the early Precambrian epochs, or to attempt to use it to date the origin of life is probably preposterous. Uncorrected extrapolations backward in time suggest that families of different ancient proteins, such as the EF-Tu/EF-G and the actin/heat shock pairs, diverged at unrealistically high values of 10 billion years ago or more, i.e., before the Earth itself was formed (Doolittle, 1992). The limited coding abilities of early RNA genomes and the absence of proof-reading refinements, combined with the relaxation of natural selection following gene duplication, may have been involved in the acceleration of the molecular clock of these genes. Accelerated amino acid substitution rates probably had important evolutionary consequences. Gene conversion

mechanisms decline as duplicated genes accumulate point mutations (Walsh, 1987) and, in general, homologous recombination is hindered by sequence divergence (Radman, 1989). Rapid rates of base-pair substitution especially exacerbated by gene duplication and differentiation may have not only limited the effciency of homogenizing mechanisms, but may have favoured rapid evolution of Archean prokaryotes. This phenomenon may be underlie the explosive metabolic evolution that took place soon after the emergence of life (Lazcano and Miller, 1994).

The recognition that major biological innovations may correlate with environmental changes has led to several integrative efforts to compare the patterns of Precambrian biological and geochemical evolution (Knoll, 1992; Margulis, 1993). Although still sketchy, the recognition of what appears to be a pattern of episodic molecular phylogenies of protists is consistent with the rapid increases in eukaryotic microbial diversity and atmospheric oxygen concentrations as inferred from the geological record (Knoll, 1992).

Prokaryotic evolution may have also followed a discontinous pattern. It is possible to distinguish three major episodes of accelerated bacterial diversification associated with (a) the rapid colonization of the early Archean Earth niches; (b) the establishment of an oxidizing environment; and (c) and the evolutionary emergence of plants and animals (Knoll and Lipps, 1993). The most intriguing of these three episodes is the oldest one, which must have been preceded by a period of intense evolutionary experimentation that after a relatively short period of time led to the establishment of the ubiquitous features of mdoern cells (i.e., genome replication, ribosome-mediated protein biosynthesis, energy-generating processes, basic metabolic enzymes, and biosynthetic pathways).

After what appears to be an explosive metabolic evolution that took place soon after the beginning of life, basic genetic processes and major molecular traits appear to have persisted essentially unchanged for more than three billions of years, and represent a prodigious case of evolutionary conservatism (Lazcano and Miller, 1994). The mechanisms involved in generating this large set of traits were extremely succesful. Following the development of oxygen-releasing photosynthesis by the ancestors of

cyanobacteria, no new essential biosynthetic traits have appeared. Later events in the evolution of the biosphere involving the emergence of new metabolic pathways (e.g., polyphenols, alkaloids, steroids, lignin, cellulose, etc.) are products of multicellularity and recombination representing only fine tunings and relatively minor additions. Recognition of this pattern raises the issue of the nature of the mechanisms responsible for the lengthy periods of evolutionary stasis during which the emergence of metabolic novelties in different prokaryotic lineages has been strongly limited.

6. FINAL REMARKS

Although in the past few years the relationship between molecular biology and microbial phylogenetics has been embittered by frequent clashes and antagonism, the development of rapidly growing sequence databanks has provided a unique view of the evolution of bacterial and eukaryotic microorganisms, and has opened new perspectives in several major fields of life sciences. Molecular evolution was originally the outcome of the wedding of molecular biology with neodarwinian theory, but it has been rapidly transformed into a field of scientific enquiry in its own right. However, its full development requires not only the development of less-expensive, more rapid macromolecular sequencing techniques and more powerful computer algorithms for constructing phylogenetic trees, but also the awareness of its non-stated assumptions and more precise definitions of its conceptual framework.

As summarized by Patterson (1988), the theoretical foundations of molecular cladistics have been based on a number of central concepts, most of which were inherited from older disciplines, such as physiology, anatomy, and neodarwinism. Homology, which is one of the key concepts in evolutionary theory, was originally used by Wolfgang Goethe, Ettiene Geoffroy Saint-Hilaire, Richard Owen, and others, to describe structural resemblance to an archetype (Donoghue, 1992). In recent years it has not only been repeatedly confused with sequence similarity (Reeck et al., 1988), but is also used to describe a wider range of possible evolutionary relationships that include species- or gene-phylogeny. In fact, some classes of homology that describe

phenomena at the molecular genetic level may have no exact equivalent in orthodox evolutionary analysis of morphological traits. One such case is paralogy, a term coined by Fitch (1970) to describe the diversification of genes following duplication events. Although some aspects of paralogy may be compared to serial homology (which is the phenomenon underlying the appearance of repeated structures in the ontogeny of some metazoa), the former corresponds to duplication events that have occured along the phylogenetic history of a sequence, and not ontogenetically, i. e., as developmental phenomena of an individual plant or animal (Patterson, 1988).

Since paralogy provides evidence of gene duplication but not of speciation events, it is the basis for infering evolutionary relationships among genes, not among species. Recognition of this distinction has led to repeated recommendations on the avoidance of paralogous sequences in phylogenetic analysis (Miyamoto and Cracraft, 1991). However, the use of paralogous duplicates in outgroup analysis for determining the evolutionary polarity of character states in universal phylogenies (Gogarten et al., 1989; Iwabe et al., 1989), has rekindled keen theoretical interest in their advantageous properties. Their use, however, does pose some risks. The naive assumption that only one paralogous duplication has taken place in the set of sequences under consideration may lead to incorrect topologies (Forterre et al., 1993). Indeed, the incorporation of genes that are the result of unrecognized multiple paralogous events in a tree algorithms may be even more insidious than the problems derived by convergent evolution and lateral gene transfer. The latter phenomena are much more easily identified at the molecular level.

The recognition that paralogous duplicates expand a monophyletic group of sequences raises a number of issues not encountered in classical evolutionary analysis. From a (classical) cladistic point of view, a character state that is found only in the outgroups is primitive (Stevens, 1980). Nonetheless, in molecular phylogenetic analysis this may not be always the case. Such rule would hold if multiple paralogous duplications have taken place, and if one (or several) of the older sequences is used as an outgroup for an unrooted tree of younger sequences. This would be the case, for instance, if a myoglobin

sequence is used to root alpha (or beta) haemoglobin trees. However, this rule would not hold if an alpha haemoglobin sequence (or a set of them) is used as an outgroup for its the beta haemoglobin tree, or viceversa.

The same is true, of course, with universal phylogenetic trees derived from elongation factors (Iwabe et al., 1989). In this case neither set is older than its homologue. In this case the reconstruction of ancestral character states from dichotomously varying paralogous genes does not comes from the analysis of the outgroup, but may be inferred from the realization that the root of the tree must have been preceded by an even older, more primitive condition in which only one copy of the gene existed prior to the paralogous duplication. Recognition of this fact is likely to play a central role in future understanding of enzyme evolution during the Archean. Although it is true that the raw material for molecular cladistics analysis is restricted to sequences derived from living organisms (or from fossil samples from which ancient preserved DNA can be retrieved) and cannot be applied to extinct groups of organisms, the construction of trees derived from archaic paralogous sequences may allow us to infer evolution prior to the earliest detectable nodes.

The flourishing of molecular techniques has led into a proliferation of sequences of isolated molecular constituents of living organisms, a storehouse of data that has already provided considerable insights into the phylogeny and the diversity of microbes. But because of its very nature, molecular cladistics separates clusters of adaptative characters into a nested hierarchical set which is expected to reflect the temporal sequence of their evolutionary acquisition. However fruitful, such approach has all the demerits of a reductionist one-trait approach to biological evolution chastised in early literature as "partial phylogeny", and since the birth of molecular phylogeny has rarely been used to attempt a truly integrative analysis of complete character complexes.

Such limitation may be overcomed in several ways, some of which are part of intellectual traditions deeply rooted in comparative biology. As Georges Cuvier contended in his 1805 *Lectures in Comparative Anatomy*, the appearance of the whole skeleton can be deduced up to a certain point by examination of a single bone. The success that Cuvier had in such anatomical

reconstructions is legendary, and was based not only in his unsurpassed knowledge and intuition, but also on what he termed the "correlation of parts", i. e., the full recognition of a functional coordination of the parts of the body of a given animal (Young, 1992). Such correlation of parts is not restricted to bones and muscles; at subcellular levels, it underlies the functional coordination among the molecular components of multigenic traits such as metabolic pathways and protein biosynthesis. As shown by the intimate relationship between the biosyntheses of valine and isoleucine, their triplet assignments, and the phylogenetic proximity of their aminoacyl-tRNA synthetases, inquiries on the early evolution of the genetic code and other basic features of living systems should be understood not only by determining the molecular phylogenies of some of their isolated components or by mathetical discussions spiced with a distinct Pythagorean flavor, but with the integrative analysis of character complexes.

But for all its foibles, the relationship between molecular biology and evolutionary theory has opened new, unsuspected avenues of intellectual exploration. Never before has such a wealth of methodological approaches and empirical data been available to the students of life's phenomena. In part because of this prosperity, systematics and evolutionary biology, two of the most broadly oriented fields of life sciences, are now in a state of intellectual agitation. The symptoms are manifold; it is possible that the traditional species concept may not apply to prokaryotes, time-cherished concepts like that of the existence of kingdoms are under fire, the origin and taxonomic position of genetic mobile elements is unknown. There is an increased awareness that the understanding of the processes underlying the generation of evolutionary novelties and the origin of ontogenic patterns cannot be restricted by classical neodarwinian explanations. We are living in the midst of hectic times in which epoch-making debates are reshaping the future of the life sciences, and the development of a more integrated molecular biology may be a never-ending story. It is said that to wish someone to live in an interesting time is one of the most terrible of all Chinese curses. Whatever the outcome of current discussions and debates, for biology the putative Oriental curse may turn out to be nothing less than an intellectual blessing.

ACKNOWLEDGMENTS

We are indebted to Dr. Lynn Margulis for her critical reading of the manuscript and many suggestions. One of us (A.L.) is grateful to Drs. Monica Riley, Jorge Llorente, and Stanley L. Miller, for several useful discussions and references. We thank Drs. F. W. Doolittle, P. Forterre, J. P. Gogarten, and their coauthors, for sharing with us their results prior to publication. A.B.B. has been supported in part by a Student Scholarship provided by the Academia de la Investigación Científica, A.C. (Mexico).

REFERENCES

- Achenbach-Richter, L., Gupta, R., Stetter., K. O., and Woese, C. R. (1987) Were the original eubacteria thermophiles? *System. Appl. Microbiol.* **9**: 34-39

- Almassy, R. J. and Dickerson, R. E. (1978) *Pseudomonas cytochrome* c551 at 2.0 Å resolution: enlargement of the cytochrome *c* family. *Proc. Natl. Acad. Sci. USA* **75**: 2674-2678

- Benachenhou-Lahfa, N., Forterre, P. and Labedan, B. (1993) Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. *J. Mol. Evol.* **36**: 335-346

- Benner, S. A., Cohen, M. A., Gonnet, G. H., Berkowitz, D. B., and Johnsson, K.P. (1993) Reading the palimpsest: contemporary biochemical data and the RNA world. In R. F. Garland and J. F. Atkins (eds), *The RNA World* (Cold Sping Harbor Laboratory Press, Cold Spring Harbor), 27-70

- Bhattacharjee, J. K. (1985) a-aminoadipate pathway for the biosynthesis of lysine in lower eukaryotes. *CRC Crit. Rev. Microbiol.* **12**: 131-151

- Bowler, P. J. (1990) *Charles Darwin: The man and his influence* (Basil Blackwell, Oxford), 250 pp.

- Broda, E. (1975) *The Evolution of the Bioenergetic Process* (Pergamon Press, Oxford), 211 pp.

- Cloud, P. E. (1968) Atmospheric and hydrospheric evolution of the primitive Earth. *Science* **160**: 729-736

- Confalonieri, F., Elie, C., Nadal, M., Bouthier de la Tour, C., Forterre, P., and Duguet, M. (1993) Reverse gyrase: a helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc. Natl. Acad. Sci. USA* **90**: 4753-4758

- Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., and Leberman, R.(1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**: 249-255

- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* Vol. 5 (National Biomedical Research Foundation, Washington, D. C.), 418 pp.

- DeLey, J. (1968) Molecular biology and bacterial phylogeny. In T. Dobzhansky, M. K. Hecht, and W. C. Steere (eds), *Evolutionary Biology* (Appleton-Century-Crofts, New York), 104-156

- Doolittle, R. F. (1992) Reconstructing history with amino acid sequences. *Protein Science* **1**: 191-200

- Doolittle, R. F. (1994) Convergent evolution: the need to be explicit. *Trends Biochem. Sci.* **19**: 15-18

- Doolittle, W. F. and Brown, J. R. (1994) Tempo, mode, the progenote and the universal root. *Proc. Natl. Acad. Sci. USA* **91**: 6721-6728

- Donoghue, M. J. (1992) Homology. In E. Fox Keller and E. A. Lloyd (eds), *Keywords in Evolutionary Biology* (Harvard University Press, Cambridge), 170-179

- Feely, D. E., Chase, D. G., Hardin, E. L., and Erlandsen, S. L. (1988) Ultrastructural evidence for the presence of bacteria, viral-like particles, and mycoplasma-like organisms associated with *Giardia* spp. *J. Protozool.* **35**: 151-158

- Fitch, W. M. (1970) Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**: 99-113

- Fitch, W. M. and Upper, K. (1987) The phylogeny of tRNA sequences provides evidence of ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **52**: 759-767

- Forterre, P., Benachenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, Ch., Labedan, B. (1993) The nature of the last universal ancestor and the root of the tree of life, still open questions. *BioSystems* **28**: 15-32

- Fotherill-Gilmore, L. A. and Michels, P. A. M. (1993) Evolution of glycolysis. *Prog. Biophys. molec. Biol.* **59**: 105-203

- García-Meza, V., González-Rodríguez, A., and Lazcano, A. (1995) Ancient paralogous duplications and the search for Archean cells. In G. R. Fleischaker, S. Colonna, and P. L. Luisi (eds), *Self-Reproduction of Supramolecular Structures: from synthetic structures to models of minimal living systems* (Klüwer, Amsterdam). (in press)

- Gogarten-Boekels, M. and Gogarten, J. P. (1994) The effects of heavy meteorite bombardment on the early evolution of life --a new look at the molecular record. *Origins of Life and Evol. Biosph.* (in press)

- Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. L., Poole, J., Date, T., Oshima, Konishi, L., Denda, K., and Yoshida, M. (1989) Evolution of the vacuolar H^+-ATPase: implications for the origin of eukayotes. *Proc. Natl. Acad. Sci. USA* **86**: 6661-6665

- Gupta, R. S. (1995) Evolution of the chaperonin families (HSP60 y HSP10, and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol. Microbiol.* **15**: 1-11

- Gupta, R. S. and Golding, G. B. (1993) Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. *J. Mol. Evol.* **37**: 573-582

- Gupta, R. S. and Singh, B. (1992) Cloning of the HSP70 gene from *Halobacterium marismortui*: relatedness of archaebacterial HSP70 to its eubacterial homologs and a model of the evolution of the HSP70 gene. *J. Bacteriol.* **174**: 4594-4605

- Harvey, R. B. (1924) Enzymes of thermal algae. *Science* **60**: 481-482

- Hilario, E. and Gogarten, J. P. (1993) Horizontal transfer of ATPase genes: the tree of life becomes a net of life. *BioSystems* **31**: 111-119

- Holm, N. G., ed., (1992) *Marine Hydrothermal Systems and the Origin of Life* (Klüwer Academic Publ., Dordrecht), 242 pp.

- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* **86**: 9355-9359

- Jensen, R. A. (1976) Enzyme recruitment in the evolution of new function. *Ann. Rev. Microbiol.* **30**: 409-425

- Joenje, H. (1989) Genetic toxycology of oxygen. *Mutat. Res.* **219**: 193-208

- Jordana, X., Chatton, B., Paz-Weisshaar, M., Buhler, J. M., Cramer, F., Ebel, J. P., and Fasiolo, F. (1987) Structure of the yeast valyl-tRNA synthetase gene (*VAS1*) and the homology of its translated amino acid sequence with*Escherichia coli* isoleucyl-tRNA synthetase. *J. Biol. Chem.* **262**: 7189-7194

- Joyce, G. F., Schwartz, A. W., Miller, S. L., and Orgel, L. E. (1987) The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl. Acad. Sci. USA* **84**: 4398-4402

- Kaine, B. P., Mehr, I. J., and Woese, C. R. (1994) The sequence, and its evolutionary implications, of a *Thermococcus celer* protein associated with transcription. *Proc. Natl. Acad. Sci. USA* **91**: 3854-3856

- Kandler, O. (1994) The early diversification of life. In S. Bengtson (ed), *Early Life on Earth: Nobel Symposium No. 84* (Columbia University Press, New York), 124-131

- Kauffman, S. A. (1993) *The Origins of Order: self-organization and selection in evolution* (Oxford University Press, New York), 709 pp.

- Keefe, A. D., Lazcano, A. and Miller, S. L. (1994) Evolution of the biosynthesis of the branched-chain amino acids. *Origins of Life and Evol. Biosph.* (in press)

- Kletzin, A. (1992) Molecular characterization of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases. *Nucleic Acid Res.* **20**: 5389-5396

- Knoll, A. H. (1992) The early evolution of eukaryotes: a geological perspective*Science* **256**: 622-627

- Knoll, A. H. and Lipps, J. H. (1993) Evolutionary history of prokaryotes and protists. In J. H. Lipps (ed), *Fossil Prokaryotes and Protists* (Blackwell Scientific Publ., Boston), 19-29

- Kolb, V. M., Dworkin, J. P., and Miller, S. L. (1994) Alternative bases in the RNA world: the prebiotic synthesis of urazole and its ribosides. *Jour. Mol. Evol.* **38**: 549-557

- Lazcano, A. (1993) Biogenesis: some like it very hot. *Science* **260**: 1154-1155

- Lazcano, A. (1994a) The RNA world, its predecessors and descendants. In S. Bengtson (ed), *Early Life on Earth: Nobel Symposium No. 84* (Columbia University Press, New York), 70-80

-Lazcano, A. (1994b) The transition from non-living to living. In S. Bengtson (ed), *Early Life on Earth: Nobel Symposium No. 84* (Columbia University Press, New York), 60-69

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

- Lazcano, A. and Miller, S. L. (1994) How long did it take for life to begin and evolve to cyanobacteria? *Jour. Mol. Evol.* **39**: 546-554

- Margulis, L. (1993) *Symbiosis in Cell Evolution*, 2nd. Ed. (W. H. Freeman Co., New York), 452 pp.

- Margulis, L. and Guerrero, R. (1991) Kingdoms in turmoil. *New Scientist* **132**: 46-50

- Mayr, E. (1990) A natural system of organisms. *Nature* **348**: 491

- Miller, S. L. and Bada, J. L. (1988) Submarine hot springs and the origin of life. *Nature* **334**: 609-611

- Miyamoto, M. M. and Cracraft, J. (1991). Phylogenetic inference, DNA sequence analysis, and the future of molecular systematics. In M. M. Miyamoto and J. Cracraft (eds), *Phylogenetic Analysis of DNA Sequences* (Oxford University Press, New York), 3-17

- Moran, N. A., Munson, M. A., Baumann, P., and Ishikawa, H. (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. Royal Soc. London B* **253**: 167-171

- Müller, M. (1988) Energy metabolism of protozoa without mitochondria. *Ann. Rev. Microbiol.* **42**: 465-488

- Nemanic, P. C., Owen, R. L., Stevens, R. P., and Mueller, J. C. (1979) Ultrastructural observations on giardiasis in a mouse model. II Endosymbiosis and organelle distribution in *Giardia muris* and *Giardia lamblia*. *Jour. Infec. Diseases* **140**: 222-228

- Nagel, G. M. and Doolittle, R. F. (1991) Evolution and relatedness in two aminoacyl-tRNA synthetase families. *Proc. Natl. Acad. Sci. USA* **88**: 8121-8125

- Norris, A. and Berg, P. (1964) Mechanism of aminoacyl RNA synthesis: studies with isolated aminoacyl adenylate complexes of isoleucyl RNA synthetase. *Proc. Natl. Acad. Sci. USA* **52**: 330-337

- Nuttall, G. H. F. (1904) *Blood Immunity and Blood Relationship: a demostration of certain blood-relationships amongst animals by means of the precipitin test for blood* (Cambridge University Press, Cambridge), 475 pp.

- Ochman, H. and Wilson, A. C. (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**: 74-86

- Oparin, A. I. (1938) *The Origin of Life* (MacMillan, New York), 270 pp

- Ouzonis, C. and Sander, C. (1992) TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell* **71**: 189-190

- Patterson, C. (1988) Homology in classical and molecular biology. *Mol. Biol. Evol.* **5**: 603-625

- Pisani, F. M., De Martino, C., and Rossi, M. (1992) A DNA polymerase from the archaeon *Sulfolobus solfataricus* shows sequence similarity to family B DNA polymerases. *Nucleic Acid Res.* **20**: 2711-2716

- Radman, M. (1989) Mismatch repair and the fidelity of genetic recombination.*Genome* **31**: 68-73

- Razin, S. (1978) The mycoplasmas. *Microbiol. Rev.* **42**: 414-470

- Reeck, G. R., de Häen, C., Teller, D. C., Doolittle, R. F., Fitch, W., Dickerson, R. E., Chambon, P., McLachlan, A. D., Margoliash, E., Jukes, T. H., and Zuckerkandl, E. (1987) "Homology" in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell* **50**: 667

- Riley, M. (1993) Functions of the gene products of *Escherichia coli. Microbiol. Rev.* **57**: 862-952

- Rivera, M. C. and Lake, J. A. (1992) Evidence that eukaryotes and eocyte prokaryotes are inmediate relatives. *Science* **257**: 74-76

- Schopf, J. W. (1993) Microfossils of the Early Archean Apex Chert: new evidence of the antiquity of life. *Science* **260**: 640-646

- Schwartz, . M. and Dayhoff, M. O. (1978) Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* **199**: 395-403

- Segerer, A. H., Burograf, S., Fiala, G., Huber, G., Huber, R., Pley, U., and Stetter, K. O. (1993) Life in hot springs and hydrothermal vents. *Origins of Life and Evol. Biosph.* **23**: 77-90

- Sidow, A. and Bowman, B. H. (1991) Molecular phylogeny. *Current Opinion Genet. Develop.* **1**: 451-456

- Simpson, G. G. (1964) The nonprevalence of humanoids. *Science* **143**: 769-775

- Sleep, N. H., Zahnle, K. J., Kastings, J. F., and Morowitz, H. J. (1989) Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature* **342**: 139-142

- Slesarev, A. I., Stetter, K. O., Lake, J. A., Gellert, M., Krah, A., Kozyavkin, S. A. (1993) DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote. *Nature* **364**: 735

- Sogin, M. L. (1994) The origin of eukaryotes and evolution into major kingdoms. In S. Bengtson (ed), *Early Life on Earth: Nobel Symposium No. 84* (Columbia University Press, New York) (in press)

- Stark, G. R. and Wahl, G. M. (1984) Gene amplification. *Ann. Rev. Biochem.* **53**: 447-491

- Stetter, K. O. (1994) The lesson of archaebacteria. In S. Bengtson (ed), *Early Life on Earth: Nobel Symposium No. 84* (Columbia University Press, New York), 114-122

- Stevens, P. F. (1980) Evolutionary polarity of character states. *Ann. Rev. Ecol. Syst.* **11**: 333-358

- Tiboni, O., Cammarano, P., and Sanangelantoni, M. A. (1993). Cloning and sequencing of the gene encoding glutamine synthase I from the archaeum *Pyrococcus woesei*: anomalous phylogenies inferred from analysis of the archaeal and bacterial glutamine synthase I sequences. *J. Bacteriol.* **175**: 2961-2969

- Wächtershäuser, G. (1990) The case for the chemoautotrophic origins of life in an iron-sulfur world. *Origins of Life Evol. Biosph.* **20**: 173-182

- Wald, G. (1954) The origin of life. *Sci. Am.* **191**: 44-53

- Wallace, D. C. and Morowitz, N. H. (1973) Genome size and evolution. *Chromosoma* **40**: 121-126

- Walsh, J. B. (1987) Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* **117**: 543-557

- Wheelis, M. L., Kandler, O., and Woese, C. R. (1992) On the nature of global classification. *Proc. Natl. Acad. Sci. USA* **89**: 2930-2934

- Woese, C. R. (1965) On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA* **54**: 1546-1552

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

- Woese, C. R. (1993) The archaea: their history and significance. In M. Kates, D. J. Kushner, and A. T. Matheson (eds), *The Biochemistry of the Archaea (Archaebacteria)* (Elsevier Science Publishers, Amsterdam), vii-xxix

- Woese, C. R. and Fox, G. E. (1977) The concept of cellular evolution. *Jour. Mol. Evol.* **10**: 1-6

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

- Wong, T. J. F. (1975) A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. USA* **72**: 1909-1912

- Yarus, M. (1993) An RNA-amino acid affinity. In R. F. Gasteland and J. F. Atkins (eds), *The RNA World* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor), 205-217

- Ycas, M. (1974) On the earlier states of the biochemical system. *J. Theor. Biol.* **44**: 145-160

- Young, D. (1992) *The Discovery of Evolution* (Natural History Museum Publications, Cambridge), 256 pp.

- Zuckerkandl, E. and Pauling, L. (1965) Molecules as documents of evolutionary history. *Jour. Theoret. Biol.* **8**: 357-366

Journal of Molecular Evolution

Emile Zuckerkandl, Editor-in-Chief
Institute of Molecular Medical Sciences
460 Page Mill Road
Palo Alto, California 94306
Phone: (415) 617-1852
Fax: (415) 322-3009

December 25, 1996

Dr. Antonio Lazcano
Unité de Physicochimie des Macromolécules biologiques
Institut Pasteur
28, rue du Docteur Roux
75724 Paris, France

Fax: 33-1-4061-3060

Mon cher Tonio,

This relates to the paper entitled "Polyphyletic gene losses can bias backtrack characterizations of the cenacestor," by Arturo Becerra, Sara Islas, José Ignacluo Leguina, Ervin Silva, and yourself.

I come back to you with one review, which may have to do -- the other is just not forthcoming, despite reminders... Please find the single review appended.

The subject of this contribution is very interesting and important right now.

Would it not seem that minimum genomes for cellular life should be established for autotrophic organisms rather than for parasites? Parasites are expected to shed a lot of genes, don't they, since they can use many substances produced by the host and can much simplify their own functions, in principle. (Morphologically this is very striking in parasitic metazoa.) I think that parasites should be left out altogether when an effort is made to establish the outlines of minimal cenancestral genomes. It is *after* the size and composition of such genomes has been derived convincingly that it will become timely to compare these genomes with those of noneukaryotic parasites, in a study of the effects of parasitism on genomes.

How about the following proposal. You and your coauthors would shorten this manuscript to the extent possible, and at the same time bolster its precise information content (two self contradictory exhortations) so that it can be considered the "Point" part of an edition of the "Point Counterpoint" rubric. Since the paper is largely built on a consideration of the contribution of particular authors (albeit to a different journal), namely, Musheglan and Koonin, a copy of your revised

manuscript would be sent to these authors, with the question whether they want to write a comment of their own (within a short time). The discussion would stop there. They need not respond, of course, and if they don't, your paper would be published under the same rubric, even though the counterpoint part would be missing.

Please let me know what you think would be best.

Warmest wishes for a happy New Year!

As ever,

Emile

Polyphyletic gene losses can bias backtrack characterizations of the cenancestor

Arturo Becerra, Sara Islas, José Ignacio Leguina,
Ervin Silva, and Antonio Lazcano*

Facultad de Ciencias
Universidad Nacional Autónoma de México
Apartado Postal 70-407
Cd. Universitaria, 04510 México, D. F.
MEXICO

FAX No. 525/616-0451
e-mail: alar@hp.fciencias.unam.mx

* corresponding author

Introduction

Mushegian and Koonin (1996) have recently published the results of a detailed comparison of the complete genomes of *Haemophilus influenzae* and *Mycoplasma genitalium* in conjuntion with the fragmentary data from other organisms available as of March 1996. Once parasite-specific sequences were discarded, the final outcome was an inventory of 256 genes which may resemble not only the genetic complement of the ancestor of Gram positive and Gram negative bacteria, but probably also the amount of DNA required today to sustain a minimal cell. Since most of these sequences have eukaryotic and/or archaeal homologs, Mushegian and Koonin discuss how this figure may be reduced to describe the genome of the last common ancestor (LCA) of the Bacteria, Archaea, and Eucarya, i.e., the cenancestor, and suggest how insights on even earlier stages of evolution can be achieved. Given the rapid pace at which more and more cellular genomes are being completely mapped and sequenced, the assumptions and strategies used in such approaches merit considerable attention. As argued here, important pitfalls can be avoided if the polyphyletic gene losses that have taken place in widely separated lineages are properly acknowledged.

The cenancestor probably had a DNA genome

The backtrack methodology proposed by Mushegian and Koonin (1996) is quite straightforward, and partly based on the idea that genes that are not found in both bacteria and eucarya, or in bacteria and archaea, were probably absent from the cenancestor. The non-stated assumption is that the archaea and eucarya are sister groups, an evolutionary relationship supported by an increasingly larger amount of molecular data. However, such approach can inadvertently miss nuclear-encoded genes which may have been part of the LCA but lost independently in both the bacterial and archeal domains, or not present in the prokaryotic genomes of a given dataset. For instance, the absence in their sample of eukaryotic or archaeal homologs of several key proteins involved in DNA replication led Mushegian and Koonin to speculate that the cenancestor may have had an RNA genome.

Several objections can be raised against this conclusion: (a) sequence similarities shared by many ancient, large proteins found in all three domains suggest that considerable fidelity already existed in the operative genetic system of their common ancestor, but such fidelity is unlikely to be found in RNA-based genetic systems; (b) sequence analysis and biochemical characterization of a ribonucleotide reductase from the archaeon *Pyrococcus furiosus* has shown that this enzyme shares considerable similarities with both its eubacterial and eukaryotic counterparts (Riera et al 1997); and (c) as underlined by Mushegian and Koonin (1996), their analysis was performed before any complete archaeal or eucaryal genomes became available in the public databases, and should thus be considered preliminary. Indeed, release of the entire *Methanococcus jannaschii* genome has allowed the identification of one archaeal DNA polymerase exhibiting sequence similarity and three conserved motifs with the eubacterial DNA polymerase II, and with the eukaryotic α, γ, and ε polymerases (Bult et al 1996). Taken together, these results suggest that DNA genomes and polymerases with proofreading and synthesizing functions evolved prior to the divergence of the three primary kingdoms.

To salvage or not to salvage

Until a more complete dataset is available, backtrack inferences on the nature of the cenancestor should be considered as preliminary and perhaps biased by the reduced genomic content of parasites, many of which have undergone multiple secondary losses For instance, the *de novo* purine nucleotide biosynthesis is probably one of the oldest metabolic pathways, but it also one of the most easily lost by a wide range of obligate symbionts and parasites. Failure to recognize such polyphyletic streamlining processes, which have taken place in *H. influenzae* and at an even greater degree in *M. genitalium*, can lead to some misunderstanding. It would be tempting, for instance, to interpret the absence of purine biosynthesis in the minimal set defined by Mushegian and Koonin (1996) as evidence that the growth and reproduction of the first lifeforms depended on the heterotrophic uptake of nucleotides present in the primitive soup (see, for instance, Pennisi 1996).

However, such conclusion would be at odds with the problems associated with the chemical synthesis and accumulation under primitive conditions not only of ribose, but also of purine- and pyrimidine ribosides, which suggest that none of them are truly prebiotic compounds (cf. Lazcano and Miller 1996).

The phylogenetic distribution of purine nucleotide salvage enzymes can also lead to some confusion regarding the cenancestor's metabolic capabilities. Based on their dataset, Mushegian and Koonin (1996) conclude that their minimal cell had the complete nucleotide salvage pathways for all bases except thymine. Adenine deaminase (ADA), which catalizes the hydrolytic deamination of adenine into hypoxanthine, is absent in both *H. influenzae* and *M. genitalium*, and therefore was not included in such inventory. However, since the ADA gene is found in other non-pathogenic Gram positive and Gram negative bacteria, it may have been part of the LCA genome. The same is probably true of the GMP reductase *guaC* gene. Since GMP reductase is not found in *H. influenzae*, *M. genitalium*, *M. jannaschii*, and *Saccharomyces cerevisiae*, it could be argued that the cenancestor lacked *guaC*. Such conclusion is not supported by the presence of GMP reductase in a group of widely separated species that includes *Escherichia coli*, *Tritrichomonas foetus*, *Trypanosoma cruzi*, *Leishmania mexicana*, and humans (Berens et al 1995). Even organisms with close phylogenetic affinities can differ in their salvage abilities. Hypoxanthine- and guanine phosphoribosyltransferase activities have been found in cell extracts of the euryarchaeota *Methanococcus voltae* (Bowen et al 1996), but the corresponding genes appear to be absent in the closely related *M. jannaschii*, where the only recognizable purine phosphoribosyltransferase gene is that of adenine PRTase (Bult et al 1996).

Molecular phylogenies are not rooted in the origin of life

The pioneering work of Mushegian and Koonin (1996) is an important improvement over previous attempts to characterize the LCA (Lazcano 1995 and references therein) but it can be improved by systematic efforts to identify streamlining processes that have led to polyphyletic gene losses in widely separated species. This may

be particularly significant given the choice of model organisms whose entire DNA is being sequenced, some of which have been selected because of their relatively small, compact genomes. It is expected that in few years larger volumes of genomic data reflecting a broader cross-section of biological diversity will become available. This will allow not only more precise descriptions of the gene complements of ancestral states, but also an understanding of the effects of parasitism on genomes and the dynamics of gene losses.

Genome sequencing and analysis is rapidly becoming a key element in our understanding of early biological evolution, but it is difficult to see how its applicability can be extended beyond a threshold that corresponds to a period of evolution in which protein biosynthesis was already in operation. Older stages are not yet amenable to this type of analysis. The first lifeforms were probably simpler than any cell now alive, and may have lacked not only familiar traits like protein catalysts, but perhaps even genetic macromolecules with ribose-phosphate backbones (Lazcano and Miller, 1996). Given the huge gap in our understanding of the evolutionary transition between the prebiotic synthesis of organic compounds and the cenacestor, the temptation to describe the nature of the very first living systems based solely on molecular cladistics should be carefully avoided.

Acknowledgements

We thank Dr. Marc Fontecave and his coworkers for providing us with their results prior to publication. The work of J. I. L. has been supported by the Consejo Superior de Investigaciones Cientificas (CSIC, Madrid, Spain). A. L. is an Affiliate of the NSCORT (NASA Specialized Center for Research and Training) in Exobiology at the University of California, San Diego.

References

Berens RL, Krug, EC, Marr, J (1995) Purine and pyrimide metabolism. In Marr JJ, Müller M (eds) *Biochemistry and Molecular Biology of Parasites* (Academic Press, London) 89-117

Bowen TL, Lin WC, Whitman WB (1996) Characterization of guanine and hypoxanthine phosphoribosyltransferase activities in *Methanococcus voltae. J Bacteriol* **178**: 2521-2556

Bult CJ, White O, Olsen, GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Relch CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geohagen NSM, Weldman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk MP, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273**: 1058-1073

Lazcano A (1995) Cellular evolution during the early Archean: what happened between the progenote and the cenancestor? *Microbiologia SEM* **11**: 1-13

Lazcano A, Miller SL (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell* **85**: 793-798

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

Pennisi E (1996) Seeking life's bare (genetic) necessities. *Science* **272**: 1098-1099

Riera J, Robb F T, Weiss R, Fontecave M (1997) Ribonucleotide reductase in the archaeon *Pyrococcus furiosus*: a critical enzyme in the evolution of DNA genomes. *Proc Natl Acad Sci USA* (in press)

Kluwer Academic Publishers
P.O. Box 17, 3300 AA Dordrecht, The Netherlands
Telephone: (0)78-334911, Telex: (0)-78-334254

Consent to Publish & Transfer of Copyright

For the mutual benefit and protection of Authors and Publishes it is necessary that Authors provide formal written *Consent to Publish* and *Transfer of Copyright* before publication of their Work.

The signed Consent ensures that the Publisher has the Author's permission to publish the relevant Work.

The signed Transfer entitles the Publisher on behalf of the Author to protect the Work against unauthorized use and to authorize dissemination of the Work by means of offprints, legitimate photocopies, microform editions, reprints, translations, and secondary information sources such as abstracting and indexing services including data bases.

The Publisher hereby requests the Author to complete and return this form promptly to the Publisher so as to ensure the proper conduct of business.

Ms. #763

Title of Contribution: "The Role of Gene Duplication in the Evolution of Purine Nucleotide Salvage Pathways"

Authors: Arturo Becerra and Antonio Lazcano

Name of Journal: Origins of Life and Evolution of the Biosphere

1. The Author hereby assigns to the Publisher the copyright to the Contribution names above whereby the Publisher shall have the exclusive right to publish the said Contribution and translations of it wholly or in part throughout the World during the full term of copyright including renewals and extensions and all subsidiary rights subject only to the following.

2. The Author retains the right to republish the Contribution in any collection consisting solely of the Authors own Works without charge and subject only to notifying the Publisher of the intent to do so and to ensuring that the publication by the Publisher is properly credited and that the relevant copyright notice if repeated verbatim.

3. In the event of receiving any other request to reprint or translate all or part of the Contribution the Publisher shall endeavor to obtain the approval of the Author prior to giving any such permission.

4. The author guarantees that the Contribution is original, has not been published previously, is not under consideration for publication elsewhere, and that any necessary permission to quote from another source has been obtained. (A copy of any such permission is supplied herewith).

N.B. Authors should delete from the foregoing any detail with which they cannot comply.

Date November 18, 1996 Name Antonio LAZCANO

Signature* A. Lazcano-Araujo

Telephone: 525/622-4823 Fax No.: 525/616-0451 or 525/622-4828

*To be signed by the Author, who hereby agrees to inform any co-authors, or to be signed by the Employer, where appropriate.

For Publisher's use only
Ms. Ref. No._____

Journal Issue _____ Pages _____ Year_____

THE ROLE OF GENE DUPLICATION IN THE EVOLUTION OF PURINE NUCLEOTIDE SALVAGE PATHWAYS

Arturo Becerra and Antonio Lazcano*

Facultad de Ciencias, UNAM
Apdo.Postal 70-407
Cd. Universitaria
04510 México, D.F., MEXICO

FAX No. 525/616-0451

e-mail: alar@hp.fciencias.unam.mx

*corresponding author

ABSTRACT

Purine nucleotides are formed *de novo* by a biochemical route that appears to be of monophyletic origin, or are synthesized from preformed purine bases and nucleosides through different salvage pathways. Phylogenetic analysis of the enzymes that participate in purine nucleotide salvage pathways provides evidence of ancient duplication events. Three monophyletic sets of enzymes, each of which catalyzes mechanistically similar reactions, can be recognized: (a) adenine-, xanthine-, hypoxanthine- and guanine- phosphoribosyltransferases, which are all homologous among themselves, as well as to nucleoside phosphorylases; (b) the set formed by adenine deaminase, adenosine deaminase, adenosine monophophate deaminase, and adenylosuccinate synthase; and (c) the pair formed by guanine reductase and inosine monophosphate dehydrogenase. These results support the idea that substrate specificity is the outcome of gene duplication, and that the purine nucleotide salvage pathways were assembled by a patchwork process that probably took place before the divergence of the three cell domains. Nucleoside and nucleotides are unstable to hydrolysis and are very unlikely prebiotic compounds. Thus, salvage pathways must have appeared after the biosynthesis of phosphorylated sugars had evolved. A simpler scheme of purine nucleotide biosynthesis based on the ability of adenine PRTase to catalyze the condensation of 4-aminoimidazole-5-carboxamide (AICA) with PRPP is also presented.

Keywords: AICA, AICAR, gene duplication, patchwork assembly, salvage pathways, purine nucleotide biosynthesis, metabolic evolution.

ABBREVIATIONS

AICA, 4-aminoimidazole-5-carboxamide; AICN, 4-aminoimidazole-5-carbonitrile; AICAR, 5-amino-4-imidazolecarboxamide ribotide; PRPP, 5-phospho-α-D-ribosyl-1-pyrophosphate; PRTase, phosphoribosyltransferase; ADA, adenine deaminase; AMP, adenosine 5′monophosphate; IMP, inosine 5′-monophosphate; GMP, guanosine 5′-monophosphate; XMP, xanthosine 5′-monophosphate; HMP, hypoxanthosine 5′-monophosphate; IMPDH, inosine monophosphate dehydrogenase; PNP, purine nucleoside phosphorylase.

1. INTRODUCTION

The first detailed attempt to explain the origin of metabolic pathways is due to Horowitz (1945), who suggested that they are the outcome of the stepwise, sequential acquisition of enzymes in reverse order as found in extant routes. Twenty years after the so-called retrograde hypothesis was first suggested, the discovery of operons prompted Horowitz (1965) to propose that clusters of genes involved in biosynthetic routes were the result of early tandem duplication events.

An alternative interpretation of role of gene duplication in the establishment of anabolic routes was developed independently by Waley (1969), Ycas (1974), and Jensen (1976). According to this other scheme, biosynthetic routes were assembled by a patchwork mechanism involving primitive enzymes of broad substrate specificity. Gene duplication and subsequent divergence of the new sequences would lead to the diversification of function and narrowing of specificity.

Since it is generally assumed that the first organisms were derived from the preformed organic compounds available in the primitive environment (Oparin, 1938), it is tempting to assume that their growth and reproduction depended on the heterotrophic uptake of nucleotides and other raw material present in the primitive soup. However, the problems associated with the prebiotic synthesis and accumulation of nucleotides argue not only against this possibility, but also against the retrograde hypothesis as the most likely explanation for the origin of the biosynthesis of purines from glycine (Keefe et al., 1995). This suggest that the purine nucleotide salvage pathways developed after the biosyntheses of phosphorylated sugar and purine nucleotides had appeared. It has been proposed that the PRTase-mediated attachment of purines to 5-phospho-a-D-ribosyl-1-pyrophosphate (PRPP) to form the corresponding 5'-nucleotides, which today is part of a salvage pathway (Figure 1), may have participated in an ancient form of purine nucleotide biosynthesis (Zubay, 1993). Here we suggest an additional possibility, based on a semi-enzymatic mechanism involving

the PRTase-mediated condensation of the prebiotic reagent 5-amino-4-imidazolecarboxamide with PRPP, to produce 5-amino-4-imidazolecarboxamide ribotide (AICAR). This hypothetical route, which is based on the prior evolution of PRPP biosynthesis, does not requires the first eight steps of the extant purine biosynthesis, all of which involve highly unstable intermediates.

2. Purine nucleotide salvage pathways

The *de novo* purine ribonucleotide biosynthesis is a highly conserved anabolic route whose phylogenetic distribution (Henderson and Paterson, 1973) strongly suggest that it may have already been present in the last common ancestor of the three cell domains. In contrast, there is wide variety of salvage pathways (Figure 1), which participate not only in the regulation of the intracellular concentration of nucleotides, but also in the reutilization of free intra- or extracellular bases and purine nucleosides (Berens et al, 1995). Fermentative degradation of salvaged purines and their nucleosides is a source of carbon, nitrogen and energy for some eubacteria (Munch-Peterson, 1983). Salvage pathways are essential for a group of distantly related organisms in which the *de novo* purine pathway has apparently been lost. This group is strictly dependent on exogenous purines and purine-nucleosides and nucleotides, and includes eubacteria such as *Haemophilus influenzae* and the mycoplasma, amitochondrial eukaryotes like *Giarda lamblia* and *Trichomonas vaginalis*, other protists like the kinetoplastida, parasitic helmiths like *Schistosoma mansoni*, and specialized animal tissue like that of brain (Tham et al., 1993; Berens et al., 1995; Tatusov et al ., 1996). It has been suggested that the *de novo* biosynthesis of purine nucleotides may have never evolved in the protists which lack this route (Hitchings,1982), but the possibility of a secondary loss due to a parasitic lifestyle appears to be more likely.

Recent sequence comparisons have show that several salvage pathway enzymes are part of larger families involved primarily in nucleotide metabolism (Mushegian and Koonin, 1994; Bork et al., 1995; Holm and Sander, 1996). However, no attempt has been made to discuss these evolutionary relationships from the viewpoint of the emergence and development of metabolic pathways, nor in terms of the prebiotic availability of the enzyme substrates. Knowledge of the biological distribution and diversity of salvages pathways is still fragmentary (Tables I and II). Nevertheless, as show in this paper, cladistic analysis of the available sequences suggests that purine nucleotide salvage metabolic pathways were shaped to a considerable extent by ancient paralogous duplications, i.e., via a patchwork mechanism.

2. MATERIAL AND METHODS

Amino acid and nucleotide sequences were extracted from the GenBank, EMBL, SWISS-PROT, and PIR databases. Additional information was retrieved from web sites which contain the entire genome sequences of *Haemophilus influenzae* (http://www.tigr.org/tdb/mdb/hidb/.html), *Mycoplasma genitalium* (http://www.tigr.org/tdb/mdb/mgdb/.html), *Methanococcus jannaschii* (http://www.tigr.org/tdb/mdb/mjdb/mjdb.html.), and *Saccharomyces cerevisiae* (http://genome-www.stanford.edu/Saccharomyces/). Sequences were compared one against every other one using the Pearson algorithm, which is part of the FASTA program. These sequences were also compared with the BLAST (Basic Local Aligment Search Tool) algorithm that is available in the web site: http://www.ncbi.nlm.nih.gov/BLAST/. This allowed the rapid identification of homologous protein sequences by focusing on regions shared by a pair of sequences in which a high density of identities is present (Pearson and Lipman, 1988; Pearson, 1990). Multiple amino acid sequence aligments were constructed using the MACAW program, which produces aligments of ungapped blocks detected by pairwise comparisons of the sequences in a given set (Schuler et al., 1991).

3. RESULTS AND DISCUSSION

a) Purine phosphoribosyltransferases

Purine phosphoribosytranferases (PRTases) are pentosyltransferases (EC 2.4.2) that catalyze the kinetically irreversible conversion of purines into ribonucleotides via a phosphorolytic reaction:

purine + 5-phospho-α-D ribosyl-1-pyrophosphate ⟶ 5'NMP+ pyrophosphate (1)

Purine phosphoribosyltransferases are part of a larger family of widely distributed enzymes which participate in the biosyntheses of histidine, tryptophan, purines, pyrimidines, and NAD, as well as in nucleotide salvage pathways. PRTases form a monophyletic group of enzymes with considerable degree of divergence, but their evolutionary relationship with other enzymes that participate in nucleotide salvages pathways such as nucleoside phosphorylases (PNPs), is still recognizable (Mushegian and Koonin, 1994).

Both PNP and purine PRTase activities have been identified in an wide range of organisms including members of the Bacteria, Archaea and the Eucarya (Table I), suggesting that the corresponding enzymes may have been already present in their last common ancestor. However, the phylogenetic distribution of PRTase genes follows a complex pattern in which independent secondary losses and duplications appear to have taken place in widely separated lineages. For instance, in *H. influenzae* the two non-contiguous sequences *HI0674* and *HI0692* encoding identical xanthine-guanine- PRTases have been identified (Fraser et. al., 1995; Casari et al., 1995). Hypoxanthine- (HPRTase) and guanine-phosphoribosyltransferase (GPRTase) biochemical activities have been detected in cell extracts of the euryarchaeota *Methanococcus voltae* (Bowen et al., 1996), but the corresponding genes are absent in the closely related *M. jannaschii,* in whose genome the only identifiable purine

phosphoribosyltranferase gene corresponds to adenine PRTase (Bult et al.,1996). No adenine PRTase activity has been found in the amitochondrial protist *Tritrichomonas foetus*. Purine PRTase activities also appear to be absent in the phylogenetically distant *Trichomonas vaginalis* and *Entoamoeba histolytica* (Berens et al., 1995), probably due to their parasitic lifestyle.

b) Adenine deaminase, adenosine deaminase, adenosine monophosphate deaminase, and adenylosuccinate synthase

Adenosine deaminase (EC 3.5.4..4) is a well-studied monomeric enzyme that catalyzes the irreversible biosynthesis of inosine by the hydrolytic deamination of adenosine (Eq .3). This reaction is mechanistically equivalent to (a) the adenine deaminase (ADA, EC 3.5.4.2) mediated synthesis of hypoxathine from adenine (Eq.2); and (b) the formation of IMP from AMP (Eq.4), catalized by adenosine monophosphate deaminase (EC 3.5.4.6). The products of these interconvertion reactions are reutilized as part of the general salvage pathway scheme (Figure 1):

$$\text{adenine} + H_2O \xrightarrow{\text{adenine deaminase}} \text{hypoxanthine} + NH_3 \qquad (2)$$

$$\text{adenosine} + H_2O \xrightarrow{\text{adenosine deaminase}} \text{inosine} + NH_3 \qquad (3)$$

$$AMP + H_2O \xrightarrow{\text{AMP deaminase}} IMP + NH_3 \qquad (4)$$

The evolutionary relatedness between adenosine deaminase and AMP deaminase had long been suspected due to their common reaction mechanisms (Frieden et al., 1980), and was confirmed by sequence analysis of eubacterial and eukaryotic

adenosine- and AMP deaminases (Chang et al., 1991). The reactions shown in Eq.3 and Eq.4 are mechanistically identical to that of Eq.2, and suggest a monophyletic origin of all three different deaminases. However, the alignment of these three enzymes shows relatively few conserved motifs due to a higher divergence of ADA (Figures 2a and 2b).

Three dimensional analysis of conversion patterns has shown that the similarities that adenosine deaminase shares with urease and phosphotriesterase (Jabri et al., 1995) can be extended to a larger set that includes not only AMP- and cytosine deaminases, but also dihydroorotases, allantoinases, hydantoinases, and imidazolonepropionases, all of which are part of the urease superfamily (Holm and Sander, 1996). Members of this superfamily are found in a wide variety of synthetic and degradative pathways, many of which involve nucleotides either as intermediates or as end-products (Holm and Sander, 1996). Thus, it is possible that they are the descendants of an unspecific protein which catalyzed a wide range of primitive biochemical reactions involving heterocyclic compounds.

It is likely that this protein superfamily includes adenosylosuccinate synthase (EC 6.3.4.4), a biosynthetic enzyme which achieves the anabolic conversion of IMP to AMP (Figure 1) by substituting the carbonyl oxygen at C6 with an amino group. The results of a database search using the BLAST algorithm do suggest a phylogenetic relationship of AMP deaminase with adenylosuccinate synthase. The sequences comparison depicted in Figure 3 shows the presence small of four conserved motifs. Although in prokaryotic genomes their genes are not adjacent, AMP deaminase and adenylosuccinate synthase catalyze sequential steps in the *de novo* biosynthesis of adenosine 5'-monophosphate.

Deaminases may be relatively latecomers in evolution. At 85 °C the half-lives of adenine and adenosine due to hydrolytic deaminations are 1.7 and 2.6 years, respectively (Frick et al.,1987). These reactions can be enhanced by inorganic

catalysts, as shown by the efficient clay-mediated hydrolytic deamination of both adenine and adenosine into hypoxanthine and inosine, respectively, under putative prebiotic conditions (Strasak and Sersen,1991). Under physiological conditions spontaneous deamination of adenine proceeds at such high rates that an specific enzymatic mechanism involving a N-glycosylase has evolved to excise IMP from DNA and reduce its mutagenic consequences (Singer and Kusmierek, 1982). This raises the possibility that deaminases and other hydrolytic enzymes (including cytosine deaminase, N-glycosylases, RNase H, and the exonuclease domain of DNA polymerase I) were not essential during the very early stages of metabolic evolution, since their high reaction rates would guarantee the availability of substrates required for the operation of these particular salvage pathways.

c) Inosine monophosphate dehydrogenase and guanosine monophosphate reductase

Inosine monophosphate dehydrogenase (IMPDH, EC 1.1.1.205) is a biosynthetic NAD-dependent dehydrogenase which converts IMP into XMP by catalyzing the formation of a carbonyl group at C2 of inosinate (Eq. 5). This reaction is followed by the guanylate synthase-mediated displacement of this carbonyl oxygen in the keto form of XMP by an amino group, to produce GMP (Figure 1). In both the *Escherichia coli* and *H. influenzae* genomes the genes encoding these two enzymes are adjacent (data not shown). This colinearity is probably due to the physiological link between their products, i.e., the requirement for the simultaneous availability of the two enzymes for the reactions sequence IMP XMP GMP to take place.

inosine 5´-monophosphate + NAD + H_2O ⟶ xanthosine 5´phosphate + NADH **(5)**

The opposite reaction, i. e., the reconvertion of GMP into IMP (Eq.6), is catalyzed by GMP reductase (EC 1.6.6.8), a NADP-dependent salvage pathway deamination enzyme:

guanosine 5´-phosphate + NADH ⟶ inosine 5´-phosphate + NH_3 + NADP (6)

Like PRTases, GMP reductase has a peculiar phylogenetic distribution that defies a simple explanation. It is absent in insects (Becker, 1974) and in rodents (Kanno et al., 1989), but in humans GMP reductase deficiencies may be lethal (Henikoff and Smith, 1989). The *guaC* gene encoding GMP reductase is present in *E.coli* (Andrew and Guest 1988), but absent in the *H. influenzae* (Fleischmann et al., 1995), *M. genitalium* (Fraser et al., 1995), *M. jannaschii* (Bult et al., 1996), and *S. cerevisiae* genomes. GMP reductase appear to be also absent in the apicomplexa, but has been detected in *Tritrichomonas foetus,* as well as in several kinetoplastida and mammals (Beck et al., 1994; Berens et al., 1995).

The evolutionary relationship between IMPH and GMP reductase was first reported by Andrews and Guest (1988). Recent databases searches have shown that both enzymes are part of a β/α-barrel protein subclass in which a conserved phosphate-binding site is found near their C-terminus (Bork et al., 1995). Availability of a larger set of IMPDH and GMP reductase sequences has confirmed previous alignments, and has allowed a further characterization of the conserved large domains (Figure 4).

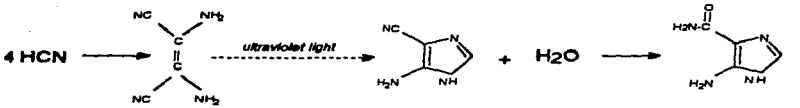
Since IMPDH and GMP reductase catalyze sequential steps in different purine metabolic pathways (Weber et al., 1992; Bork et al., 1995), it could be argued that their monophyletic origin raises the possibility that they may be a particular case of the retrograde hypothesis. Functional contiguity of homologous enzymes, however, does not probes by itself the Horowitz hypothesis (Fani et al.,1995), since alternative

explanations based on a primitive, less specific enzyme catalyzing both reactions, are equally likely. This possibility is consistent with the homology of IMPDH and GMP reductase with other β/α barrel enzymes, many of which use heterocyclic substrates (Bork et al., 1995). Due to its hydrolytic instability, it is probable that GMP was absent from the primitive soup. Thus, it is likely that GMP reductase appeared after its biosynthetic homologue IMPDH.

4. The semi-enzymatic origin of purine nucleotide biosynthesis: a new hypothesis

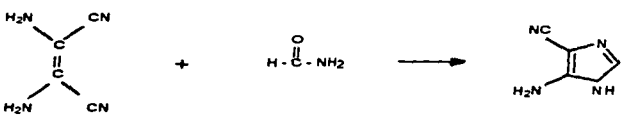
PRPP and other phosphorylated sugars are unlikely components of the primitive soup (Larralde et al., 1995). Thus, purine PRTase activity could have not appeared when life originated but only after riboside-synthesizing metabolic routes had evolved. The bioynthesis of purines from glycine pathway cannot be based on components from the primitive soup, since all intermediates are ribosides which are unstable to hydrolysis. Furthermore, ribose is generally considered not to have been a significant component of the prebiotic environment (Shapiro, 1988; Larralde et al., 1995).

On the other hand, it is generally agreed that purines and their prebiotic precursors were present in the primitive environment. This includes 5-amino-4-imidazolecarboxamide (AICA), which is a key intermediate in a potentially prebiotic synthesis of guanine and hypoxanthine (Sanchez et al., 1968). Two prebiotic routes to AICA can be envisoned. Sunlight irradiation of the HCN tetramer, diaminomaleonitrile produces 4-aminoimidazole-5-carbonitrile in relatively good yelds (Ferris and Orgel, 1966) (Eq. 7).

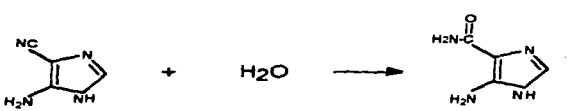


diaminomaleonitrile 4-aminoimidazole-5-carbonitrile 4-aminoimidazole-5-carboxamide (7)

Under basic conditions diaminomaleonitrile reacts with formamide and also produces 4-aminoimidazole-5-carbonitrile (AICN) (Eq. 8), whose hydrolysis also produces AICA (Eq. 9) in good yields of ~20% (Oró and Kimball, 1962; Lowe et al., 1963; Sanchez et al., 1968):



diaminomalonitrile formamidine 4-aminoimidazole-5-carbonitrile (8)



4-aminoimidazole-5-carbonitrile 4-aminoimidazole-5-carboxamide (9)

AICA is not known to play any direct role in extant organisms. Quite surprisingly, it is an alternate substrate for adenine PRTase, an enzyme which not only catalyzes the formation of AMP from PRPP and adenine (Eq. 1), but also the direct condensation of PRPP with the prebiotic reagent AICA, to yield 5-amino-4-imidazolecarboxamide ribotide AICAR (K_{eq}=9.7, pH 8) (Flaks et al., 1957), which is an intermediate in the *de novo* biosynthesis of purine nucleotides (Figure 1). AICA can be converted into hypoxanthine in high yields (40%) in concentrated ammonium formate (Zubay, 1993). In the scheme suggested here, AICAR would undergo an equivalent non-enzymatic

formylation to produce 4-carboxamide-5-formamidoimidazole ribotide. A dehydration reaction would follow, leading to inosine-5'-monophosphate. As shown in Figure 5, a simpler, primitive semi-enzymatic route of purine nucleotide biosynthesis may thus be envisoned:

Our proposal implies that adenine PRTase was originally a key catalyst in a primitive version of the anabolic route of purines. This hypothetical route is supported by the likelihood that importans amounts of HCN and its derivatives were available in the primitive Earth due to eutectic freezing (Miller and Orgel, 1974), or because their formation was enhanced by different prebiotic catalysts such as glyconitrile (Schwartz and Goverde, 1982). Our proposal is also supported by the chemical stability of AICAR (Flaks et al., 1957), which is 114 days at 100 °C and pH 7, and 108 days at 100 °C and pH 8 (Sanchez et al., 1968). No data on the half-lives of AICAR are have been stimated for lower temperatures, but AICAR is approximately a hundred-times more stable than its corresponding nitrile, AICN, whose half-life at 0 °C pH 8 is 2×10^3 years (Sanchez et al., 1968) suggesting that is was indeed available for the scheme suggested here to take place. We provide no explanation for the emergence of PRPP biosynthesis, and assume that the origin of phosphorylated sugar metabolisms is clearly related to the emergence of the ribose-phosphate backbone of RNA. This remains an open question, since the nature of the backbone of the first genetic polymer is unknown (Lazcano and Miller, 1996).

5. CONCLUSIONS

Enzymes that participate in purine metabolism exhibit relatively good levels of 35 to 65% of sequence similarity between distantly related organisms (Johnson et al., 1987; Henikoff and Smith, 1989). However, as discussed in this paper, the components of purine nucleotide salvage pathways have complex phylogenetic histories. Availability of several complete cellular genomes has allowed insights into the versatility of salvage strategies followed by different organisms (Table II), but whether a simple underlying pattern exists or not requires information that is still not within reach. For instance, while PRTase and nucleoside phosphorylase activities appear to be essential, and have not been lost in *H. influenzae* and *M. genitalium*, adenine deaminase is clearly disposable (Table II).

It is likely that the major salvage pathways were established during early stages of metabolic evolution, perhaps prior to the divergence of the three main cell lines (Mushegian y Koonin, 1996a). Unlike other macromolecules such as 16S rRNA, the peculiarities of the phylogenetic distribution of the salvage pathway enzymes show that these are not good universal molecular markers. Nevertheless, they are a good model for the study of metabolic evolution. Direct uptake of purines (Hitchings, 1982) and pyrimidines from the prebiotic soup by the first organisms may be considered as primitive form of salvage pathway, but as argued here, the available evidence indicates that the pathways involving PRTases, deaminases, GMP reductases, and other enzymes that use nucleosides and nucleotides as substrates appeared only after their biosynthetic counterparts had already evolved. However ancient, extant salvage pathways of purine nucleotides did not appear when life originated, i. e., they are not truly primordial.

Since comparative genomic analysis has demonstrated that gene order is not conserved in prokaryotic evolution (Mushegian and Koonin, 1996b; St. Jean and

Charlebois, 1996; Tatusov et al., 1996), lack of contiguity of homologous genes encoding salvage pathway enzymes (data no shown) does not disproofs by itself the Horowitz hypothesis. However, lack of support for the retrograde hypothesis is provided by the cladistic analysis of the enzymatic sequences of salvage pathway. With the exception of the pairs formed by (a) IMPDH and GMP reductase, and (b) AMP deaminase and adenylosuccinate synthase, no other case was found in which sequential enzymatic steps are catalyzed by homologous enzymes. Neither of these two examples can be considered definitive proof of the retrograde hypothesis, since alternative explanations based on the existence of primitive less-specific enzymes are equally plausible and are in fact sopported by the homology of salvage enzymes with many other proteins.

The discovery that a significant portion of bacterial genomes is the outcome of ancient paralogous duplications (Fleischmann et al., 1995, Fraser et al., 1995; Koonin et al., 1995; Labedan and Riley, 1995) is consistent with the hypothesis of the patchwork assembly of metabolic routes, and may be invoked to explain the evolution of salvage pathways. Each of the sets of homologous enzymes discussed here is formed by proteins that catalyze mechanistically similar reactions. This supports the idea that ancient pathways were mediated by enzymes of broad specifity (Waley, 1969; Ycas, 1974; Jensen, 1976), which may have participated in metabolic routes that today are not directly connected, such as the histidine and pyrimidine biosynthetic pathways.

Recognition of the role of gene duplication in metabolic evolution does not answers questions related to the emergence and number of the original starter types, i.e., of the enzymes that did not arise in this manner. In some cases, the starter types may stem from slow non-enzymatic reactions were the protein improved on a previously sluggish process (Lazcano and Miller, 1996), as in the case of the photochemical decarboxylation of orotic acid which yields uracil described by Ferris and Joshi (1979). Primitive pathways may have existed in which only few steps were mediated by enzymes. In other cases, semi-enzymatic syntheses may have taken place (Miller and

Lazcano, in prep.). One example may be the model of purine nucleotide biosynthesis presented here, in which participate both prebiotic reagents like AICA, and biological catalysts such as adenine PRTase. In this regard, it is interesting to note that several enzymes including nitrogenase (Silver and Postgate, 1973), urease (Estermair et al., 1992), and adenine PRTase (Flaks et al., 1957), catalyze reactions involving HCN and/or its derivatives (cyanides, acetylene, cyanamide, dicyanamide, AICA), all of which may be prebiotic reagents. Since these compounds rarely participate in contemporary biochemical process, their use as alternate substrates by different enzymes raises the possibility that these are vestigial activities from a time in which metabolism depended on semi-enzymatic processes (Miller and Lazcano, in prep.).

ACKNOWLEDGMENTS

This paper was presented at the Orleans ISSOL'96 Metting, July 5-12, 1996. We thank Stanley L. Miller for calling to our attention to the paper by Estermaier et al. (1972), and Chris Sander for sharing with us his results prior to publication. Economic support from the ISSOL Travel Fund is gratefully acknowledged (A.B.) This paper was completed during a leave of absence of one of us (A. L.) as Visiting Professor at the Institut Pasteur (Paris), during which he enjoyed the hospitality of Professor Henri Buc and his associates at the Unité de Physicochemie des Macromolécules Biologiques. Support from the Maulio Cantarini Foundation is gratefully acknowledged. A.L. is an Affiliate of the NSCORT (NASA Specialized Center for Research and Training) in Exobiolgy at the University of California, San Diego.

REFERENCES

Andrews, S.C. and Guest J.R. : 1988, *Bichem. J.* **255**, 35-43.

Beck, J. T., Zhao, S., Wang, C. C.: 1994, *Exp. Parasitol.* **78**, 101-112

Becker, J.-L.: 1974, *Biochemie* **56**, 1249-1253

Berens, R. L ., Krug, E. C., and Marr, J.J.: 1995, Marr, J.J. and Müller, M. (eds), *Biochemistry and Molecular Biology of Parasites*, Academic Press, London, 89-117

Bork, P., Gellerich, J., Groth, H., Hooft, R., and Martin, F.: 1995, *Protein Science* **4**: 268-274

Bowen, T.L.,Lin, W.C.,and Whitman, W.B.: 1996, *J. Bacteriol.* **178**, 2521-2526

Bult, C.J., White, O., Olsen, G.J.,Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F.,Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, M.P., Fraser, C.M., Smith, H.O., Woese, C.R., and Venter, J.C.: 1996, *Science* **273**, 1058-1073

Casari, G., Andrade, M.A., Bork, P., Boyle, S.J., Ouzounis, C., chneider, R., Tamames, J., Valencia, A., and Sander, C.:1995, *Nature* **376**, 647-648

Chang, Z., Nygaard, P., Chinault, A.C., and Kellems, R.E.: 1991, *Biochemistry* **30**, 2273-2280

Estermaier, L.M., Sieber, A.H., Lottspeich, F., Matern, D.H.M. and Hartmann, G.R. (1992). *Angew. Chem. Int. Ed. Engl.* **31**, 620-622

Fani, R., Lio, P., and Lazcano, A. : 1995: *J. Mol. Evol.* **41**, 760-774

Ferris, J. P. and Joshi, P. C.: 1979, *J. Org. Chem.* **44**, 2133-2137

Ferris, J. P. and Orgel, L. E.: 1966, *J. Am. Chem.* Soc. **88**, 1074

Flaks, J.G., Erwin, M.J., and Buchanan, J.M.: 1957, *J. Biol. Chem.* **228**, 201-213

Fleischmann, R.D., Adams, M.D.,White, O., Clayton, R.A., Kirkness, E.F.,Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G.,

FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C.: 1995, *Science* **269**, 496-512

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M. Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson III, C.A., and Venter, J.C.: 1995, *Science* **270**, 397-403

Frick, L., MacNeela, J.P., and Wolfenden, R.: 1987, *Bioorganic Chem.* **15**, 100-108

Frieden, C., Kurz, L.C., and Gilbert, H.R.: 1980, *Biochemistry* **19**, 5303-5309

Henderson, J.F., and Paterson, R.P.: 1973, *Nucleotide Metabolism: an introduction*, Academic Press, New York

Henikoff, S. and Smith, J. M.: 1989, *Cell* **58**, 1021-1022

Hitchings, G.H.: 1982, *Adv. Enzyme Regul.* **20**, 375-386

Holm, L., and Sander, C.: 1996, *Proteins* (in press)

Horowitz, N. H.: 1945, *Proc. Natl. Acad. Sci. USA* **31**, 153-157

Horowitz, N.H.: 1965, In Bryson, V., and Vogel, H.J. (eds), *Evolving Genes and Proteins*, Academic Press, New York, 15-23

Jabri, E., Carr, M.B., Hausinger, R.P., and Karplus, P.A.: 1995, *Science* **268**, 998-1004

Jensen, R.A.: 1976, *Annu. Rev. Microbiol.* **30**, 409-425

Johnson, D. H., Edstrom, J. E., Burnett, J. B., and Friedman, T. B.: 1987, *Gene* **59**, 77-86

Kanno, H., Huang, I. Y., Kan, Y. W., and Yoshida, A.: 1989, *Cell* **58**, 595-606

Keefe, A.D., Lazcano, A., and Miller, S.L.: 1995, *Origins Life Evol. Biosph.* **25**: 99-110

Koonin, E.V., Tatusov, R.L., and Rudd, K.E.: 1995, *Proc. Natl. Acad. Sci. USA* **92**, 11921-11925

Labedan, B. and Riley, M.: 1995, *Mol. Biol. Evol.* **12**, 980-987

Larralde, R., Robertson, M.P., and Miller, S.L.: 1995, *Proc. Natl. Acad. Sci. USA* **92**, 8158-8160

Lazcano, A. and Miller, S.L.: 1996, *Cell* **85**, 793-798

Lowe, C.U., Rees, M.W., and Markham, R.: 1963, *Nature* **199**, 219-222

Munch-Petersen, A (ed): 1983, *Metabolism of Nucleotides, Nucleosides, and Nucleobases in Microorganisms*, Academic Press, London

Mushegian, A.R. and Koonin, E.V.: 1994, *Protein Science* **3**,1081-1088

Mushegian, A.R. and Koonin, E. V.: 1996a, *Proc. Natl. Acad. Sci. USA* **93**, 10268-10273

Mushegian, A.R. and Koonin, E.V.: 1996b, *TIGS* **12**, 289-290

Oparin, A.I.: 1938, *The Origin of Life*, Macmillan, New York

Oró, J. and Kimball, A.P.: 1962, *Arch. Biochem. Biophys.* **96**,293-313

Pearson, W.R.: 1990, in Doolittle, R.F. (ed), In *Methods in Enzymology* **138**, 63-98

Pearson, W.R. and Lipman, D.I.: 1988, *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448

Sanchez, R.A., Ferris, J.P., and Orgel, L.E.: 1968, *J. Mol. Biol.* **38**, 121-128

Schuler, G.D., Altschul, S.F., and Lipman, D.J.: 1991, *Proteins Struct. Funct. Genet.* **9**:180-190

Schwartz, A. W. and Goverde, M.: 1982, *J. Mol. Evol.* **18**, 351-353

Shapiro, R.: 1988, *Origins Life Evol. Biosph.* **18**, 71-85

Shapiro, R.: 1995, *Origins Life Evol. Biosph.* **25**, 83-98

Silver, W.S. and Postgate, J.R.: 1973, *J. Theoret. Biol.* **40**, 1-10

Singer, B. and Kusmierek, J.T.: 1982, *Ann. Rev. Biochem.* **52**, 655-693

St. Jean, A. and Charlebois, R. L.:1996, *J.Bacteriol.* **178**, 3860-3868

Strasak, M; and Sersen, F.: 1991, *Naturwissenschaften* **78**, 121-122

Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, M.S., Borodovsky, M., Rudd, K.E., and Koonin, E.V.: 1996, *Current Biol.* **6**, 279-291

Tham, T.N., Ferris, S., Kovacic, R., Montagnier, L., and Blanchard, A.: 1993, *J. Bacteriol.* 175, 5281-5285

Waley, S.G.: 1969, *Comp. Biochem. Physiol.* **30**, 1-7

Weber, G., Nakamura, H., Natsumeda, Y., Szekeres, T., and Nagai, M.: 1992, *Adv. Enzyme Regul.* **32**, 57-69

Ycas, M.: 1974, *J. Theoret. Biol.* **44**, 145-160

Zubay, G.: 1993, *Chemtracts - Biochem. Mol. Biol.* **4**, 317-323

FIGURE CAPTIONS

Figure 1. Purine nucleotide metabolism. Both the *de novo* and the major salvage pathways routes are shown.

Figure 2.a Aligment of the deduced *E. coli* (Ec ADA), *Mus musculus* (Mm ADA), and *Homo sapiens* (Hs ADA) adenosine deaminase amino acid sequences, with the *S. cerevisiae* (Sc AMD) and *H. sapiens* (Hs AMD) AMP deaminases. Stars (*) indicate identical amino acids, and (.) similar ones.

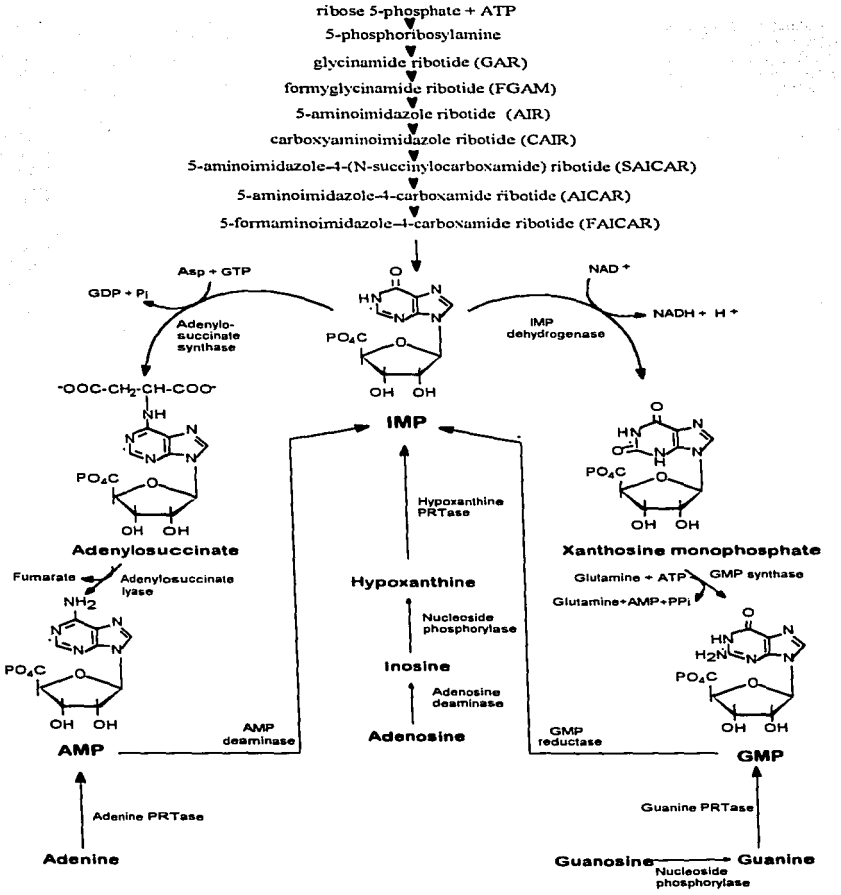
Figure 2.b Aligment of the deduced *E. coli* (Ec ADA), *Mus musculus* (Mm ADA), and *H. sapiens* (Hs ADA) adenosine deaminase amino acid sequences, with the *E. coli* (Ec ADEC), *Bacillus subtilis* (Bs ADEC), and *M. jannaschii* (Mj ADEC), adenine deaminases. Stars (*) indicate identical amino acids, and (.) similar ones.

Figure 3. Aligment of the deduced *B. subtilis* (Bs PURA) and *Dictyostelium discideum* (Dd PURA) adenylosuccinate synthase amino acid sequences, with the deduced *H. sapiens* (Hs AMD) and *Ratus norvegicus* (Rn AMD) AMP deaminase amino acid sequences. Stars (*) indicate identical amino acids, and (.) similar ones.

Figure 4. Aligment of the deduced *Arabidopsis thaliana* (At IMPD), *Acinetobacter calcoaceticus* (Ca IMPD), *B.subtilis* (Bs IMPD), *Drosophila melanogaster* (Dm IMPD), *E.coli* (Ec IMPD), *Leishmania donavan* (Ld IMPD), *Mesocricetus auratus* (Ma IMPD), *S. cerevisiae* (Sc IMPD), and *Pyrococcus furiosus* (Pf IMPD) IMP dehydrogenase amino acid sequences, with the *E. coli* (Ec GUAC), *Ascaris lumbricoides* (Al GUAC), and *H. sapiens* (Hs GUAC) GMP reductase deduced

amino acid sequences. Stars (*) indicate identical amino acids, and (.) similar ones.

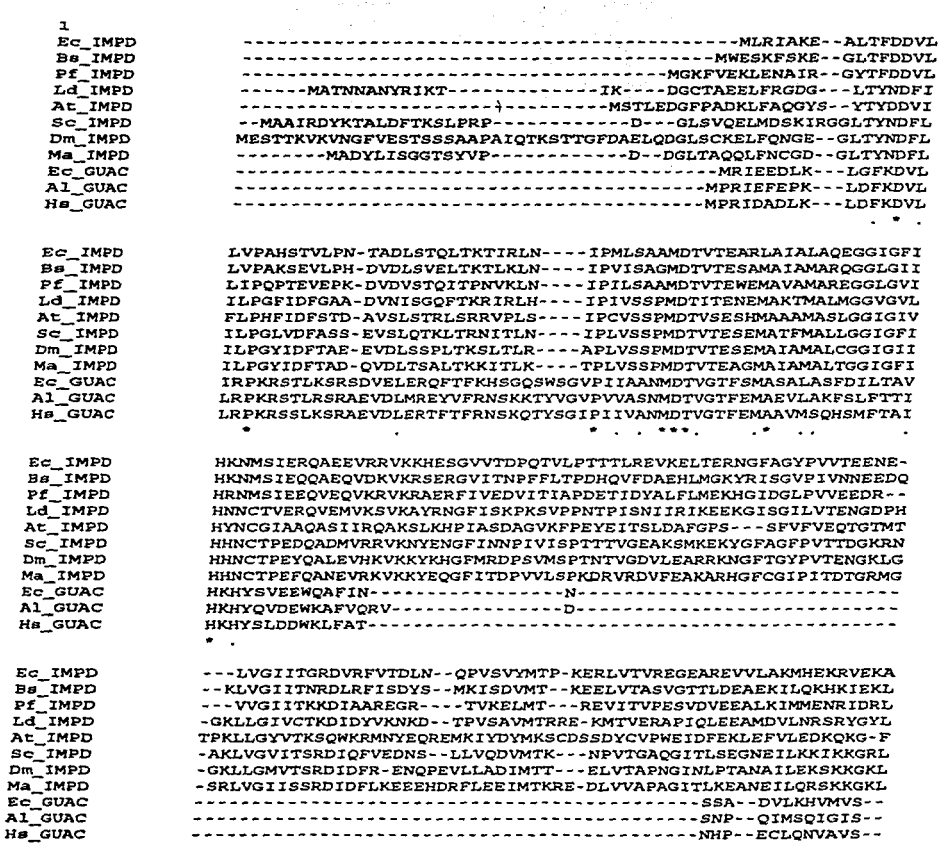
Figure 5. A possible ancestral semi-enzymatic AICA-dependent biosynthesis of purine nucleotides


ribose 5-phosphate + ATP
5-phosphoribosylamine
glycinamide ribotide (GAR)
formylglycinamide ribotide (FGAM)
5-aminoimidazole ribotide (AIR)
carboxyaminoimidazole ribotide (CAIR)
5-aminoimidazole-4-(N-succinylocarboxamide) ribotide (SAICAR)
5-aminoimidazole-4-carboxamide ribotide (AICAR)
5-formaminoimidazole-4-carboxamide ribotide (FAICAR)
Asp + GTP
GDP + Pi
Adenylo-succinate synthase
NAD+
NADH + H+
IMP dehydrogenase
IMP
Adenylosuccinate
Fumarate
Adenylosuccinate lyase
Hypoxanthine PRTase
Hypoxanthine
Xanthosine monophosphate
Glutamine + ATP
GMP synthase
Glutamine+AMP+PPi
Nucleoside phosphorylase
Inosine
Adenosine deaminase
Adenosine
AMP deaminase
AMP
GMP reductase
GMP
Adenine PRTase
Adenine
Guanine PRTase
Guanosine
Nucleoside phosphorylase
Guanine

```
360
Ec ADA        IHRHLDGNIRPQTILELGR---QY----------NISLPAQSLETLIPHVQVIANEPDLV
Mm ADA        LHVHLDGAIKPETILYFGK---KR----------GIALPADTVEELR-NIIGMDKPLSLP
Hs ADA        LHVHLDGSIKPETILYYGR---RR----------GIALPANTAEGLL-NVIGMDKPLTLP
Sc AMD        THVHHSACMNQKHLLRFIKHKLRHSKDEKVIFRDGKLLTLDEVFRSL-HLTGYDLSIDTL
Hs AMD        THIHASSCMNQKHLLRFIKRAMKRHLEEIVHVEQGREQTLREVFESM-NLTAYDLSVDTL
               * *    .     .*   .    .          .                ..

Ec ADA        SFLT------KLD----WGVKVLAS--LDACR----RVAFENIEDAARHGLHYVE-LRFS
Mm ADA        GFLA------KFD----YYMPVIAG-CREAIK----RIAYEFVEMKAKEGVVYVE-VRYS
Hs ADA        DFLA------KFD----YYMPAIAG-CREAIK----RIAYEFVEMKAKEGVVYVE-VRYS
Sc AMD        DMHAHKDTFHRFDKFNLKYNPIGESRLREIFLKTNNYIKGTYLADITKQVIFDLENSKYQ
Hs AMD        DVHADRNTFHRFDKFNAKYNPIGESVLREIFIKTDNRVSGKYFAHIIKEVMSDLEESKYQ
                 .    . *           .      .          .   .  .*  ..

Ec ADA        PGYMAMAHQLPV-------------AGVVEAVIDGVREGCRTFGVQAKLIG------IMS
Mm ADA        PHLLANSKVDPM--PWNQTEGDVTPDDVVDLVNQGLQEGEQAFGIKVRS-----I--LCC
Hs ADA        PHLLANSKVEPI--PWNQAEGDLTPDEVVALVGQGLQEGERDFGVKARS-----I--LCC
Sc AMD        NCEYRISVYGRSLDEWDKLASWVIDNKVISHNVRWLVQIPRLYDIYKKTGIVQSFQDICK
Hs AMD        NAELRLSIYGRSRDEWDKLARWAVMHRVHSPNVRWLVQVPRLFDVYRTKGQLANFQEMLE
                  .                   *       . .  . . .            .

Ec ADA        RTFEAACQQELEAF---LAHR--DQITALDLAGDELG-----------FPG--------
Mm ADA        MRHQPSWSLEVLEL---CKKYNQKTVVAMDLAGDETI------------EG--SS-LFP-
Hs ADA        MRHQPNWSPKVVEL---CKKYQQQTVVAIDLAGDETI------------PG--SS-LLP-
Sc AMD        NLFQPLFEVTKNPQSHPKLHVFLQRVIGCDSVDDESK-VDRRFHRKYPKPSLWEAPQNPP
Hs AMD        NIFLPLFEATVHPASHPELHLFLEHVDGFDSVDDESKPENHVFNLESPLPEAWVEEDNPP
                          .    .     *    **

Ec ADA        ---------SLFLSHFNRARDAGWH-ITV--HAGEAAGPESIWQAIRELGAERIGHGVKA
Mm ADA        ---------GHVEAY-EGAVKNGIH-RTV--HAGEVGSPEVVREAVDILKTERVGHGYHT
Hs ADA        ---------GHVQAY-QEAVKSGIH-RTV--HAGEVGSAEVVKEAVDILKTERLGHGYHT
Sc AMD        YSYYLYYLYSNVASLNQWRAKRGFNTLVLRPHCGEAGDPE--HLVSAYLLAHGISHGILL
Hs AMD        YAYYLYYTFANMAMLNHLRRQRGFHTFVLRPHCGEAGPIH--HLVSAFMLAENISHGLLL
                                 *  .   .  *.**                .  .  . **

Ec ADA        IEDRALMDFLAEQQIGIESCLTSNIQTSTVAELAAHPLKTFLEHGIRASINTDDPGVQGV
Mm ADA        IEDEALYNRLLKENMHFEVCPWSSYLTGAWDPKTTHAVVRFKNDKANYSLNTDDPLIFKS
Hs ADA        LEDQALYNRLRQENMHFEICPWSSYLTGAWKPDTEHAVIRLKNDQANYSLNTDDPLIFKS
Sc AMD        RKVPFVQYLYYLDQVGIAMSPLSN--NALFLTYDKNPFPRYFKRGLNVSLSTDDPLQFSY
Hs AMD        RKAPVLQYLYYLAQIGIAMSPLSN--NSLFLSYHRNPLPEYLSRGLMVSLSTDDPLQFHF
                   .       ..   .  *   .         .              *. ****

Ec ADA        D---IIHEYTVAAPAAGLSREQIRQAQINGLEMAFLSAEEKRALREKVAAK--------
Mm ADA        T---LDTDYQMTKKDMGFTEEEFKRLNINAAKSSFLPEEEKKELLERLYREY-------
Hs ADA        T---LDTDYQMTKRDMGFTEEEFKRLNINAAKSSFLPEDEKRELLDLLYKAYGMPPSASA
Sc AMD        TREPLIEEYSVAAQIYKLSNVDMCELARNSVLQSGWEAQIKKHWIGKDFDKSGVEGNDVV
Hs AMD        TKEPLMEEYSIATQVWKLSSCDMCELARNSVLMSGFSHKVKSHWLGPNYTKEGPEGNDIR
                 .  .*  ..      .          *     .      *
                                                                   780
Ec ADA        ----------------------------------------------
Mm ADA        ----Q-----------------------------------------
Hs ADA        GQNL------------------------------------------
Sc AMD        RTNVPDIRINYRYDTLSTELELVNHFAN--FKRTIEEK----------
Hs AMD        RTNVPDIRVGYRYETLCQELALITQAVQSEMLETIPEEAGITMSPGPQ
```

```
 60
Ec ADA          --------------------MIDT---T-LP-LTDIHRHLDGN-IRP----QTILELGRQ
Mm ADA          --------------------MAQTP-AFNKP-KVELHVHLDGA-IKP----ETILYFGKK
Hs ADA          --------------------MAQTP-AFDKP-KVELHVHLDGS-IKP----ETILYYGRR
Mj ADEC         FVDLNDEIDKIIEKIKEDVKVIDLKGKYLSPTFIDGHIHIESSHLIPSEFEKFVLKSGVS
Ec ADEC         GVGAEYTDAP-------ALQRIDARGATAVPGFIDAHLHIESSMMTPVTFETATLPRGLT
Bs ADEC         GLGEYEGEN-----------IIDAEGQMIVPGFIDGHVHIESSMVTPIEFAKAVLPHGVT
                                                *  . * *..  . *      *  *

Ec ADA          YNISLPAQSLETLIP-HVQVIANEP--DLVSFLTKLDWGVKVLAS-----LDACRRVAFE
Mm ADA          RGIALPADTVEELR--NIIGMDKPL--SLPGFLAKFDYYMPVIAGC----REAIKRIAYE
Hs ADA          RGIALPANTAEGLL--NVIGMDKPL--TLPDFLAKFDYYMPAIAGC----REAIKRIAYE
Mj ADEC         KVVIDPHEIANIAGKEGILFMLNDA--KILDVYVMLPSCVPATNLETS-GAEITAENIEE
Ec ADEC         TVICDPHEIVNVMGEAGPAWFARCAEQARQNQYLQVSSCVPALEGCDVNGASFTLEQMLA
Bs ADEC         TVVTDPHEIANVSGEKGIEFMLEQARHTPLNIHFMLPSSVPAASFERS-GAILKAADLKP
                  .  *                                .

Ec ADA          NIEDAARHGL----HY-------------VELRFSPGYMAMAHQLPVAG-----------
Mm ADA          FVEMKAKEGV----VY-------------VEVRYSPHLLANSKVDPMPWNQTEGDVTP--
Hs ADA          FVEMKAKEGV----VY-------------VEVRYSPHLLANSKVEPIPWNQAEGDLTP--
Mj ADEC         LILLDNVLGLGEVMNYPAVINEDEEMLKKIEVAKKYNKLIDGHCPKLKGWELNKYISHGI
Ec ADEC         WRDHPQVTGLAEMMDYPGVISGQN--ALLDKLDAFRHLTLDGHCPGLGGKELNAYITAGI
Bs ADEC         FYEEEEVLGLAEVMDYVSVQQAEKDMVQKLLDARVGGKRIDGHLAGLSTDLINIYRTAFV
                          *.    *                    .   .

Ec ADA          -----------VVEAVIDGV----REGCRTFGVQAKLIGIMSRTFEAACQQELEAFLAHR
Mm ADA          ---------DDVVDLVNQGL----QEGEQAFGIKVRSILCCMRHQPSWSLEVLELCKKYN
Hs ADA          ---------DEVVALVGQGL----QEGERDFGVKARSILCCMRHQPNWSPKVVELCKKYQ
Mj ADEC         MSDHESVDEDEALEKLRLGLKLMIREGTASKNIYLLNICKKIKDFRNIMLVSDDVCIKDL
Ec ADEC         ENCHESYQLEEGRRKLQLGMSLMIREGSAARNLNALAPLINEFNSPQCMLCTDDRNPWEI
Bs ADEC         LNDHEVTSKEEALYRIRRGMYVMMREGSVAKNTLNVLPAVNEKNARRFFFCTDDKHVDDL
                              .  *.    .**    .                .
```

```
1
Bs Pura   -----------------MSSVVVVGTQWGDEG-KGKITDFLSENAEVIARYQGGNNAGH
Dd Pura   ------------------MASIIIGSQWGDEG-KGKLVDILSQQFDVVARCQGGANAGH
Rn AMD    MPLFKLTGQGKQIDDAMRSFAEKVFASEVKDEGGRHEISPFDVDEICPISLREMQAHIFH
Hs AMD    MPLFKLPAEEKQIDDAMRNFAEKVFASEVKDEGGRQEISPFDVDEICPISHHEMQAHIFH
                           .  .  ..  ***  .  .          ..  .  .  *

Bs Pura   ----TIKFD--------GITYKLHLIP----SGIFYKDKTCVIGNG-MVVDPKAL-VTEL
Dd Pura   ----TIVVD--------GKKIALHLIP----SGILNEKASCILGNG-MVIHLPTF-FKEV
Rn AMD    MENLSMSMDGRRKRRFQGRKTVNLSIPQSETSSTKLSHIEEFISSSPTYESVPDFQRVQI
Hs AMD    LETLSTSTEARRKKRFQGRKTVNLSIPLSETSSTKLSHIDEYISSSPTYQTVPDFQRVQI
            .   .          *       **     *         .              ..

Bs Pura   AYLHERNVSTDN--LRISN--RAHVILP-YHLK-LDEVEEERKG----AN-KIGTTKKGI
Dd Pura   QGLQDKGINYKGR-LFVSD--RAHLVFD-LHQM-IDAMKEAELSNGTSND-SIGTTKRGI
Rn AMD    TGDYASGVTVEDFEVVCKGLYRALCIREKYMQKSFQRFPKTPSKYLRNIDGEALVAIESF
Hs AMD    TGDYASGVTVEDFEIVCKGLYRALCIREKYMQKSFQRFPKTPSKYLRNIDGEAWVANESF
               ..     .       **    .                          .

Bs Pura   GPAYMDKAARI--GIRIADLLD--RDAFAEK--LERNLEEKNRL--LE-KMYETEGFK-L
Dd Pura   GPCYSSKASRG--GLRVCDLYS--PEHFRKT--FTRLVENKH-------KRFGSFEYD-V
Rn AMD    YPVFTPPPKKGEDPFRREDLPANLGYHLKMKGGVIYIYPDEAAASRDEPKPYPYPNLDDF
Hs AMD    YPVFTPPVKKGEDPFRTDNLPENLGYHLKMKDGVVYVYPNEAAVSKDEPKPLPYPNLDTF
           * .    .      *   *                              *

Bs Pura   EDILDEYYEYGQQ--IKKYVCDTSVVLNDALDEGRRVLFEGAQGVMLDI----DQG-TYP
Dd Pura   EAELKRYQEFAEM--LKPFVIDSVYYLNQAFKDGKKVLIEGAQSTMLDL----DFG-CYP
Rn AMD    LDDMNFLLALIAQGPVKTYTHRRLKFLSSKF-QVHQMLNEMDELKELKNNPHRDFYNCRK
Hs AMD    LDDMNFLLALIAQGPVKTYTHRRLKFLSSKF-QVHQMLNEMDELKELKNNPHRDFYNCRK
    .          .*  .       *       ...* *  .    *     *   .

Bs Pura   FVTSSNPVAAVSQSVLVSARPKIKHVVGVSKAYTTRVGDGPFPTELKDEIGDQIREVGRE
Dd Pura   YVTSSASSVGGACTGLGISPNKVVTQIGVVKAYTTKVGSGPFPTEQNDHVGDSLRKAGSE
Rn AMD    VDTHIHAAACMNQKHLLRFIKKSYHIDADRVVYSTKEKNLTLKELFAQLNMHPYDLTVDS
Hs AMD    VDTHIHAAACMNQKHLLRFIKKSYQIDADRVVYSTKEKNLTLKELFAKLKMHPYDLTVDS
   *          *       *           *.*.

Bs Pura   YGTTTGRP--PRVGWF-DSVVVRHARRVSGITDLSLNSID------VLAGIETLKICVRY
Dd Pura   FGTTTGRP--RRIGWL-DAVVLRYTSMINDFTRLNLTKLD------VLSDFEEIKIGVDY
Rn AMD    LDVHAGRQTFQRFDKFNDKYNPVGASELRDLYLKTDNYINGEYFATIIKEVGADLVDAKY
Hs AMD    LDVHAGRQTFQRFDKFNDKYNPVGASELRDLYLKTDNYINGEYFATIIKEVGADLVEAKY
    .**      *       *       .  .        .  .      ..      .   *

Bs Pura   RYKGEIIEEF---P---ASLKALAECEPVYEEMPGWTEDITGAKSLSELPENARHYLERV
Dd Pura   KYKGETIKSF---P---ASLETLAQCEVVYESFPGWKCDLSHVTEYDQLPIQAKNYIKRI
Rn AMD    QHAEPRLSIYGRSPDEWSKLSSWFVGNRIYCPNMTWMIQVPRIYDVFRSKNFLPHFGKML
Hs AMD    QHAEPRLSIYGRSPDEWSKLSSWFVCNRIHCPNMTWMIQVPRIYDVFRSKNFLPHFGKML
.     .   .    *    .*.      .        *     .              ..  .
                                                             540
Bs Pura   SQLTGIPLSIFSVGPD-RSQTNVLRSVYRAN-----------------------------
Dd Pura   EELVGVPIVYIGVGVE-RKNLIERKELI--------------------------------
Rn AMD    ENIF-LPVFEATINPQTHPDLSVFLKHITGFDSVDDESKHSGHMFSSKSPKPEEWTMENN
Hs AMD    ENIF-MPVFEATINPQADPELSVFLKHITGFDSVDDESKHSGHMFSSKSPKPQEWTLEKN
   .   .*.     ..
```


1
Ec_IMPD
Bs_IMPD
Pf_IMPD
Ld_IMPD
At_IMPD
Sc_IMPD
Dm_IMPD
Ma_IMPD
Ec_GUAC
Al_GUAC
Hs_GUAC

```
1
Ec_IMPD          ---------------------------------------------MLRIAKE--ALTFDDVL
Bs_IMPD          ----------------------------------------------MWESKFSKE--GLTFDDVL
Pf_IMPD          ---------------------------------------MGKFVEKLENAIR--GYTFDDVL
Ld_IMPD          ------MATNNANYRIKT-------------IK----DGCTAEELFRGDG---LTYNDFI
At_IMPD          ----------------------)---------MSTLEDGFPADKLFAQGYS--YTYDDVI
Sc_IMPD          --MAAIRDYKTALDFTKSLPRP-----------D---GLSVQELMDSKIRGGLTYNDFL
Dm_IMPD          MESTTKVKVNGFVESTSSSAAPAIQTKSTTGFDAELQDGLSCKELFQNGE--GLTYNDFL
Ma_IMPD          --------MADYLISGGTSYVP-----------D--DGLTAQQLFNCGD--GLTYNDFL
Ec_GUAC          ---------------------------------------------MRIEEDLK---LGFKDVL
Al_GUAC          ---------------------------------------------MPRIEFEPK---LDFKDVL
Hs_GUAC          ---------------------------------------------MPRIDADLK---LDFKDVL
                                                                          . * .

Ec_IMPD          LVPAHSTVLPN-TADLSTQLTKTIRLN----IPMLSAAMDTVTEARLAIALAQEGGIGFI
Bs_IMPD          LVPAKSEVLPH-DVDLSVELTKTLKLN----IPVISAGMDTVTESAMAIAMARQGGLGII
Pf_IMPD          LIPQPTEVEPK-DVDVSTQITPNVKLN----IPILSAAMDTVTEWEMAVAMAREGGLGVI
Ld_IMPD          ILPGFIDFGAA-DVNISGQFTKRIRLH----IPIVSSPMDTITENEMAKTMALMGGVGVL
At_IMPD          FLPHFIDFSTD-AVSLSTRLSRRVPLS----IPCVSSPMDTVSESHMAAAMASLGGIGIV
Sc_IMPD          ILPGLVDFASS-EVSLQTKLTRNITLN----IPLVSSPMDTVTESEMATFMALLGGIGFI
Dm_IMPD          ILPGYIDFTAE-EVDLSSPLTKSLTLR----APLVSSPMDTVTESEMAIAMALCGGIGII
Ma_IMPD          ILPGYIDFTAD-QVDLTSALTKKITLK----TPLVSSPMDTVTEAGMAIAMALTGGIGFI
Ec_GUAC          IRPKRSTLKSRSDVELERQFTFKHSGQSWSGVPIIAANMDTVGTFSMASALASFDILTAV
Al_GUAC          LRPKRSTLRSRAEVDLMREYVFRNSKKTYVGVPVVASNMDTVGTFEMAEVLAKFSLFTTI
Hs_GUAC          LRPKRSSLKSRAEVDLERTFTFRNSKQTYSGIPIIVANMDTVGTFEMAAVMSQHSMFTAI
                   *                    .  . ***.   .*  ..          .

Ec_IMPD          HKNMSIERQAEEVRRVKKHESGVVTDPQTVLPTTTLREVKELTERNGFAGYPVVTEENE-
Bs_IMPD          HKNMSIEQQAEQVDKVKRSERGVITNPFFLTPDHQVFDAEHLMGKYRISGVPIVNNEEDQ
Pf_IMPD          HRNMSIEEQVEQVKRVKRAERFIVEDVITIAPDETIDYALFLMEKHGIDGLPVVEEDR--
Ld_IMPD          HNNCTVERQVEMVKSVKAYRNGFISKPKSVPPNTPISNIIRIKEEKGISGILVTENGDPH
At_IMPD          HYNCGIAAQASIIRQAKSLKHPIASDAGVKFPEYEITSLDAFGPS---SFVFVEQTGTMT
Sc_IMPD          HHNCTPEDQADMVRRVKNYENGFINNPIVISPTTTVGEAKSMKEKYGFAGFPVTTDGKRN
Dm_IMPD          HHNCTPEYQALEVHKVKKYKHGFMRDPSVMSPTNTVGDVLEARRKNGFTGYPVTENGKLG
Ma_IMPD          HHNCTPEFQANEVRKVKKYEQGFITDPVVLSPKDRVRDVFEAKARHGFCGIPITDTGRMG
Ec_GUAC          HKHYSVEEWQAFIN----------------N----------------------------
Al_GUAC          HKHYQVDEWKAFVQRV--------------D----------------------------
Hs_GUAC          HKHYSLDDWKLFAT-----------------------------------------------
                 * .

Ec_IMPD          ---LVGIITGRDVRFVTDLN--QPVSVYMTP-KERLVTVREGEAREVVLAKMHEKRVEKA
Bs_IMPD          --KLVGIITNRDLRFISDYS--MKISDVMT--KEELVTASVGTTLDEAEKILQKHKIEKL
Pf_IMPD          ---VVGIITKKDIAAREGR----TVKELMT---REVITVPESVDVEEALKIMMENRIDRL
Ld_IMPD          -GKLLGIVCTKDIDYVKNKD--TPVSAVMTRRE-KMTVERAPIQLEEAMDVLNRSRYGYL
At_IMPD          TPKLLGYVTKSQWKRMNYEQREMKIYDYMKSCDSSDYCVPWEIDFEKLEFVLEDKQKG-F
Sc_IMPD          -AKLVGVITSRDIQFVEDNS--LLVQDVMTK---NPVTGAQGITLSEGNEILKKIKKGRL
Dm_IMPD          -GKLLGMVTSRDIDFR-ENQPEVLLADIMTT---ELVTAPNGINLPTANAILEKSKKGKL
Ma_IMPD          -SRLVGIISSRDIDFLKEEEHDRFLEEIMTKRE-DLVVAPAGITLKEANEILQRSKKGKL
Ec_GUAC          ---------------------------------------SSA--DVLKHVMVS--
Al_GUAC          ---------------------------------------SNP--QIMSQIGIS--
Hs_GUAC          ---------------------------------------NHP--ECLQNVAVS--
                                                                  .
```

```
Ec_IMPD     LVVDDEFHLIGMITVKDFQKAEAKPNACKDE---QGRLRVGAAVGAGAGNEERVDALVAA
Bs_IMPD     PLVDDQNKLKGLITIKDIEKVIEFPNSSKDI---HGRLIVGAAVGVTGDTMTRVKKLVEA
Pf_IMPD     PVVNEDGKLVGLITMSDLVARKKYKNAVRNE---KGELLVAAAVSPFD--LRRAIELDRA
Ld_IMPD     PIVNENDEVVNLCSRRDAVRARDYPHSTLDK---SGRLICAAATSTRPEDKRRVAALADV
At_IMPD     VVLERDGETVNVVTKDDIQRVKGYPKSGPGTVGPDGEWMVGAAIGTRESDKERLEHLVNV
Sc_IMPD     LVVDEKGNLVSMLSRTDLMKNQNYPLASKSAN--TKQLLCGASIGTMDADKERLRLLVKA
Dm_IMPD     PIVNQAGELVAMIARTDLKKARSYPNASKDS---NKQLLVGAAIGTRSEDKARLALLVAN
Ma_IMPD     PIVNENDELVAIIARTDLKKNRDYPLASKDA---KKQLLCGAAIGTHEDDKYRLDLLALA
Ec_GUAC     -------------------------------------TG---TSDADFEKTKQILDLNP
Al_GUAC     -------------------------------------SG---ISTSDFDKLRTVCDMIP
Hs_GUAC     -------------------------------------SG---SGQNDLEKMTSILEAVP

Ec_IMPD     GVDVLLIDSSHGHSEGVLQRIRETRAKYPDLQIIGGNVATAAGARALAEAGCSAVKVGIG
Bs_IMPD     NVDVIVIDTAHGHSQGVLNTVTKIRETYPELNIIAGNVATAEATRALIEAGADVVKVGIG
Pf_IMPD     GVDVIVVDTAHAHNLKAIKAMKEMRQKVSADFIVG-NIANPKAVDDLTFA--DAVKVGIG
Ld_IMPD     GVDVLVLDSSQGNTIYQIAFIKWVKSTYPHLEVVAGNVVTQDQAKNLIDAGADGIRIGMG
At_IMPD     GVNAVVLDSSQGNSIYQLEMIKYVKKTYPELDVIGGNVVTMYQAQNLIQAGVDGLRVGMG
Sc_IMPD     GLDVVILDSSQGNSIFELNMLKWVKESFPGLEVIAGNVVTREQAANLIAAGADGLRIGMG
Dm_IMPD     GVDVIILDSSQGNSVYQVEMIKYIKETYPELQVIGGNVVTRAQAKNLIDAGVDGLRVGMG
Ma_IMPD     GVDVVVLDSSQGNSIFQINMIKYMKEKYPNLQVIGGNVVTAAQAKNLIDAGVDALRVGMG
Ec_GUAC     ALNFVCIDVANGYSEHFVQFVAKAREAWPTKTICAGNVVTGEMCEELILSGADIVKVGIG
Al_GUAC     ELEYICLDVANGYSEVFVDFIRRVREQFPTHTIFAGNVVTGEMVEELILSGADVVKVGIG
Hs_GUAC     QVKFICLDVANGYSEHFVEFVKLVRAKFPEHTIMAGNVVTGEMVEELILSGADIIKVGVG
             .  ..* .       .  .   .       .   *. .      *  .    ...*.*

Ec_IMPD     PGSICTTRIVTGVGVPQITAVADAVEALEGTGIPVIADGGIRFSGDIAKAIAAGASAVMV
Bs_IMPD     PGSICTTRVVAGVGVPQITAIYDCATEARKHGKTIIADGGIKFSGDITKALAAGGHAVML
Pf_IMPD     PGSICTTRIVAGVGVPQITAIAMVADRAQEYGLYVIADGGIKYSGDIVKAIAAGADAVML
Ld_IMPD     SGSICITQEVLACGRPQGTAVYKVAQYCASRGVPCTADGGLRQVGDICKALAIGANCAML
At_IMPD     SGSICTTQEVCAVGRGQATAVYKVCSIAAQSGIPVIADGGISNSGHIVKALVLGASTVMM
Sc_IMPD     TGSICITQEVMACGRPQGTAVYNVCEFANQFGVPCMADGGVQNIGHITKALALGSSTVMM
Dm_IMPD     SGSICITQEVMACGCPQATAVYQVSTYARQFGVPVIADGGIQSIGHIVKAIALGASAVMM
Ma_IMPD     CGSICITQEVLACGRPQATAVYKVSEYARRFGVPVIADGGIQNVGHIAKALALGASTVMM
Ec_GUAC     PGSVCTTRVKTGVGYPQLSAVIECADAAHGLGGMIVSDGGCTTPGDVAKAFAR-ADFVML
Al_GUAC     PGSVCTTRKKAGVGYPQLSAVLECADASHGLNGHVMSDGGCTNPGDVAKAFGGGADFVMI
Hs_GUAC     PGSVCTTRTKTGVGYPQLSAVIECADSAHGLKGHIISDGGCTCPGDVAKAFGAGADFVML
             **.* *.     *  * .*.               .***    * . **        *.

Ec_IMPD     GSMLAGTEESPGEIELYQGRSYKSYRGMGSLGAMSK------GSSDRYFQSDNA-ADKLV
Bs_IMPD     GSLLAGTSESPGETEIYQGRRFKVYRGMGSVAAMEK------GSKDRYFQEE---NKKFV
Pf_IMPD     GNLLAGTKEAPGKEVIINGRKYKQYRGMGSLGAMMK------GGAERYYQGGYMKTRKFV
Ld_IMPD     GGMLSGTTETPGEYFFKGGVRLKVYRGMGSLEAMS----QGKESGKRYLSEN---EAVQV
At_IMPD     GSFLAGSTEAPGGYEYTNGKRIKKYRGMGSLEAMTK------GSDQRYLGDQ---TKLKI
Sc_IMPD     GGMLAGTTESPGEYFYQDGKRLKAYRGMGSIDAMQKTGTKGNASTSRYFSES---DSVLV
Dm_IMPD     GSLLAGTSEAPGEYFFSDGVRLKKYRGMGSLEAMERGDAKGAAMSRYYHNEM---DKMKV
Ma_IMPD     GSLLAATTEAPGEYFFSDGIRLKKYRGMGSLDAMD----KHLSSQNRYFSEA---DKIKV
Ec_GUAC     GGMLAGHEESGGRIVEENGEKFMLFYGMSSESAMKR----HVGGVAEYRAAE-----GKT
Al_GUAC     GGLLAGHDQCGGEVVEKDGKKYKLFYGMSSDTAMKK----YQGSVAEYRASE-----GKT
Hs_GUAC     GGMFSGHTECAGEVFERNGRKLKLFYGMSSDTAMNK----HAGGVAEYRASE-----GKT
            *   .   .. *      *     . ** *  **             *
```

```
Ec_IMPD          PEGIEGRVAYKGRLKEIIHQQMGGLRSCMGLTGCGTIDELRTKA-----EFVRISGAGIQ
Bs_IMPD          PEGIEGRTPYKGPVEETVYQLVGGLRSGMGYCGSKDLRALREEA-----QFIRMTGAGLR
Pf_IMPD          PEGVEGVVPYRGTVSEVLYQLVGGLKAGMGYVGARNIKELKEKG-----EFVIITSAGLR
Ld_IMPD          AQGVSGNVVDKGSAAKLIAYVSKGLQQSAQDIGEISFDAIREKMYAGQVLFSRRSPTAQG
At_IMPD          AQGVVGAVADKGSVLKLIPYTMHAVKQGFQDLGASSLQSAHGLLRSNILRLEARTGAAQV
Sc_IMPD          AQGVSGAVVDKGSIKKFIPYLYNGLQHSCQDIGCRSLTLLKNNVQRGKVRFEFRTASAQL
Dm_IMPD          AQGVSGSIVDKGSVLRYLPYLECGLQHSCQDIGANSINKLRDMIYNGQLRFMKRTHSAQL
Ma_IMPD          AQGVSGAVQDKGSIHKFVPYLIAGIQHSCQDIGAKSLTQVRAMMYSGELKFEKRTSSAQV
Ec_GUAC          VK-----LPLRGPVENTARDILGGLRSACTYVGASRLKELTKRT-----TFIRVQ---EQ
Al_GUAC          IY-----MPYRGDVSRTIHDLLGGLRSACTYIGATKLKELSKRA-----TFVRVT---QQ
Hs_GUAC          VE-----VPYKGDVENTILDILGGLRSTCTYVGAAKLKELSRRA-----TFIRVT---QQ
                           .*          ..         *

Ec_IMPD          ESHVHDVTITKESPNYRLG-------------------------S-
Bs_IMPD          ESHPHDVQITVHRNKALPGLFGSHQKKTGFVYDECCQSGFFSSD--
Pf_IMPD          ESHPHDIIITNEAPNYPLER------------------------
Ld_IMPD          EGGVHSLHSYEKKLFAAKM-------------------------
At_IMPD          EGGVHGLVSYEKKSF-----------------------------
Sc_IMPD          EGGVHNLHSYEKRLHN----------------------------
Dm_IMPD          EGNVHGLFSYEKRL------------------------------F
Ma_IMPD          EGGVHSLHSYEKRLF-----------------------------
Ec_GUAC          ENRIFNNL-----------------------------------
Al_GUAC          TNDQYSAYEVPRID-----------------------------
Hs_GUAC          HNTVFS-------------------------------------
```

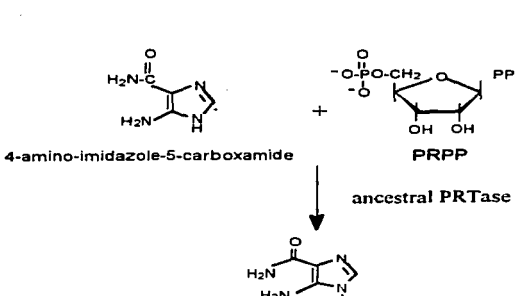

4-amino-imidazole-5-carboxamide
PRPP
ancestral PRTase
Ribose-P
AICAR
AMP and GMP

Table I. Biological distribution of PRTases and nucleoside phosphorylases

ORGANISM	APRTase	XPRTase	HGPRTase	PNP
Escherichia coli	*	*	*	*
Haemophilus influenzae	*	*	*	*
Salmonella typhimurum	*	*		?
Lactococcus lactis	*		*	?
Mycoplasma genitalium	*		*	*
Methanococcus voltae	*	*	*	?
Methanococcus jannaschii	*			*
Giardia lamblia	*			
Tritrichomonas foetus	*			*
Kinetoplastida	*	*	*	*
Leishmania donavani	*		*	?
Plasmodium falciparum	*		*	*
Toxoplasma gondii	*	*	*	*
Eimeria tenella	*			*
Schistosoma mansoni	*		*	*
mammalian cells	*		*	*

APRTase: adenine phosphoribosyltransferase; GPRTase: guanine phosphoribosyltransferase; XPRTase: xanthine phosphoribosyltransferase; HGPRTase: hypoxanthine-guanine phosphoribosyltransferase; HGXPRTase: hypoxanthine-guanine-xanthine phosphoribosyltransferase; PNP: nucleoside phosphorylase. GPRTase has only been reported in *G. lamblia*, and HGXPRTase only in *T. foetus* and *E. tenella* (Based on Berens et al., 1995).

* presence

? no information available

Table II. Purine nucleotide salvage pathways in organisms whose entire genomes have been sequenced (October, 1996)

Organism	Enzyme	Salvage pathway
Haemophilus influenzae	PRTase	direct conversion of base into a ribonucleotide
	nucleoside phosphorylase	reversible conversion of bases to nucleosides
Mycoplasma genitalium	PRTase	direct conversion of base into a ribonucleotide
	nucleoside phosphorylase	reversible conversion of bases to nucleosides
Methanococcus jannaschii	PRTase	direct conversion of base into a ribonucleotide
	nucleoside phosphorylase	reversible conversion of bases to nucleosides
	adenine deaminase	interconversion by base alterations
Saccharomyces cerevisiae	PRTase	direct conversion of base into a ribonucleotide
	nucleoside phosphorylase*	reversible conversion of bases to nucleosides
	adenine deaminase	interconversion by base alterations

* The pirS48560 sequence from *S. cerevisiae* has a 50.7% identity value in 274 aa in human PNP

CONCLUSIONES

Si bien los caracteres moleculares como el 16/18S rRNA han demostrado el origen monofilético de todos los organismos y, por ende, la existencia de un ancestro común a todos ellos (Woese 1983), es importante reconocer los limites de este enfoque tanto en la reconstrucción de las filogenias y las características del cenancestro, como respecto a la poca o nula información que nos proporcionan sobre el origen de la vida mismo. No hacerlo puede llevar a conclusiones prematuras, como ocurrió con la definición original del progenote sugerida por Woese y Fox (1977), o con la caracterización del cenancestro como una célula con genoma de RNA y que supuestamente dependía heterotróficamente de nucleótidos de origen abiótico, tal como lo sugirieron Mushegian y Koonin (1996).

El trabajo de Mushegian y Koonin (1996) que propone al cenancestro como una célula dotada de genoma de RNA, enfrenta una serie de objeciones que incluyen: a) las evidencias de gran cantidad de genes homólogos comunes a los tres dominios celulares, lo que sugiere que el cenancestro poseía un complejo sistema genético (Lazcano et al 1992); b) la homología encontrada entre las secuencias de DNA polimerasas de *Methanococcus jannaschii* con la DNA polimerasa II eubacteriana y con las polimerasas α,γ, y ε de eucariontes (Bult *et al* 1996); y c) la similitud que existe entre las secuencias de la ribonucleótido reductasa de la arqueobacteria *Pyrococcus furiosus* con sus contrapartes eubacteriana y eucarionte (Riera *et al.*, 1997). Estas evidencias sugieren que los genomas celulares de DNA surgieron antes de la divergencia de los tres dominios celulares. De hecho, la metodología de Mushegian y

Koonin (1996), basada al menos en parte en las ideas de Benner *et al* (1989), que pretende identificar los genes posteriores al *cenancestro* y a su vez reconstruir algunas de las características de los estados anteriores al mismo, ha pasado por alto la posibilidad de pérdidas secundarias polifiléticas posteriores a la diversificación de los tres linajes celulares.

Como ya se discutió arriba, cuarenta años antes de que se reconociera el papel del RNA en la evolución biológica temprana N.H. Horowitz (1945) sugirió la llamada hipótesis retrógrada para explicar el origen de las rutas biosintéticas. Los resultados presentados aquí sugieren que la aplicabilidad de esta idea es bastante limitada. El análisis cladístico de las enzimas que participan en las rutas de salvamento de nucleótidos de purinas presentadas en este trabajo no apoya la idea de Horowitz (1945, 1965), que supone que los genes homólogos codifican enzimas que catalizan pasos sucesivos en una misma ruta biosintética. El análisis comparativo de genomas celulares completos ha demostrado que el orden de los genes no se encuentra conservado en la evolución de los procariontes (Mushegian y Koonin, 1996b; St. Jean y Charlebois, 1996; Tatusov *et al.*, 1996). Sin embargo, aunque la falta de contigüidad de los genes homólogos que codifican enzimas de rutas de salvamento no descalifica por sí misma la hipótesis de Horowitz, los resultados del análisis comparativo de las secuencias de estas enzimas sí son una prueba en contra. Con la excepción de los pares formados por (a) la IMPDH y la GMP reductasa; y (b) la AMP deaminasa con la adenilosuccinato sintasa, no existen ejemplos adicionales que demuestren que enzimas homólogas catalizan pasos sucesivos. Estos ejemplos son pocos y no bastan, por sí mismos, para validar la hipótesis retrograda. Se puede suponer, por ejemplo, que los pasos son químicamente equivalentes y que la enzima

original era menos específica. En este caso, la evidencia de la homología de las enzimas mencionadas podría ser interpretada como apoyo a las ideas de Waley (1969), Ycas (1974) y Jensen (1976).

Los componentes enzimáticos de las rutas de salvamento de nucleótidos de purinas presentan una historia filogenética compleja. El acceso a genomas celulares completos ha permitido demostrar que las rutas de salvamento presentan una variedad de estrategias en diferentes organismos, aunque no es claro aún si existe un patrón o no en su distribución. A diferencia de macromoléculas como el RNA ribosomal, la distribución filogenética de las rutas de salvamento demuestra que éstas no son buenos marcadores evolutivos de tipo universal. Sin embargo, constituyen un buen modelo para estudiar el origen y la evolución de las rutas metabólicas. El uso directo de purinas y pirimidinas presentes en la sopa primitiva por los primeros organismos, puede considerarse como una forma primitiva de ruta de salvamento (Hitchings, 1982), pero es probable que las rutas de salvamento que dependen de PRTasas, deaminasas, GMP reductasas y otras enzimas que utilizan como substrato a nucleósidos y nucleótidos hayan surgido luego de su contraparte biosintética. Es decir, el origen de la mayoría de las rutas de salvamento de nucleótidos de purinas no es concomitante con el origen de la vida mismo.

El descubrimiento de que una porción importante de los genomas bacterianos ha resultado de duplicaciones parálogas ancestrales (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995; Koonin *et al.*, 1995; Labedan y Riley, 1995) es consistente con la hipótesis de que las rutas metabólicas fueron ensambladas por *patchwork*. Cada uno de los ejemplos de enzimas homólogas que se discuten en este trabajo son proteínas que catalizan reacciones químicas comparables. Ello

apoya la idea de que las rutas ancestrales fueron mediadas por enzimas que poseían una baja especificidad al substrato (Waley, 1969; Ycas, 1974; Jensen, 1976), que pudieron participar en rutas metabólicas que actualmente no se encuentran directamente conectadas, como la biosíntesis de histidina y pirimidinas. Por otro lado, la demostración del papel que la duplicación génica jugó en el ensamblaje de las rutas de salvamento de nucleótidos de purinas no solamente es consistente con la hipótesis de *patchwork*, sino que también permite explicar la aparente rapidez con la que pudieron haber evolucionado las rutas anabólicas durante el Arqueano temprano (Lazcano y Miller, 1994).

El reconocimiento del papel de la duplicación génica en la evolución de las rutas metabólicas no resuelve el problema del origen las enzimas que no surgieron de esta manera. En algunos casos, las enzimas ancestrales pudieron haber tenido su origen en reacciones no enzimáticas, y fueron proteínas que aceleraron un proceso espontáneo pero lento (Lazcano y Miller, 1996), como es el caso de la descarboxilación fotoquímica del ácido orótico descrita por Ferris y Joshi (1979). Todo sugiere que las rutas primitivas pudieron haber existido con un número reducido de pasos mediados por enzimas poco específicas. Un ejemplo lo constituye el modelo de síntesis de purinas propuesto en el tercer artículo (Becerra y Lazcano, 1997), en donde participan tanto reactivos prebióticos como el AICA y catalizadores biológicos como la adenina PRTasa. En este sentido, es interesante hacer notar que varias enzimas, incluyendo las nitrogenasas (Silver y Postgate, 1973), las ureasas (Estermair *et al.*, 1992) y la adenina PRTasa (Flacks *et al.*, 1957), catalizan reacciones que involucran HCN o sus derivados (cianida, acetileno, cianamida, dicianamida, AICA, etc.) los cuales son considerados como reactivos prebióticos. El hecho de que estos compuestos casi no participen en los

procesos bioquímicos actuales, pero que se usen como substratos alternos por diferentes enzimas, abre la posibilidad de que estemos observando vestigios de cuando los metabolismos primitivos dependían de procesos semi-enzimaticos.

REFERENCIAS

Becerra, A. and Lazcano, A.:1997, *Origins Life Evol. Biosphere* (enviado)

Benner, S., A.:1989, *Proc Natl Acad Sci U S A* **86**, 7054-7058

Berens, R. L ., Krug, E. C., and Marr, J.J.: 1995, Marr, J.J. and Müller, M. (eds), *Biochemistry and Molecular Biology of Parasites*, Academic Press, London, 89-117

Bork, P., Gellerich, J., Groth, H., Hooft, R., and Martin, F.: 1995, *Protein Science* **4**: 268-274

Bult, C.J., White, O., Olsen, G.J.,Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Relch, C.I., Overbeek, R., Kirkness, E.F.,Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, M.P., Fraser, C.M., Smith, H.O., Woese, C.R., and Venter, J.C.: 1996, *Science* **273**, 1058-1073

Estermaier, L.M., Sieber, A.H., Lottspeich, F., Matern, D.H.M., and Hartmann, G.R. (1992). *Angew. Chem. Int. Ed. Engl.* **31**, 620-622

Fani, R., Lio, P., and Lazcano, A. : 1995: *J. Mol. Evol.* **41**, 760-774

Fitch, W.M., and Upper, K. M., *Cold Spring Harbor Symp. Quant. Biol.* **52**,759-767

Flaks, J.G., Erwin, M.J., and Buchanan, J.M.: 1957, *J. Biol. Chem.* **228**, 201-213

Fleischmann, R.D., Adams, M.D.,White, O., Clayton, R.A., Kirkness, E.F.,Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C.: 1995, *Science* **269**, 496-512

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M. Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson III, C.A., and Venter, J.C.: 1995, *Science* **270**, 397-403

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, J., Date, T., Oshima, T., Konishi,L., Denda, K., and Yoshida,M.:1989, *Proc. Natl. Acad. Sci. USA* **86**, 6661-6665

Gogarten, J. P.:1994, *J. Mol. Evol. 39*, 541-543

Henderson, J.F. and Paterson, R.P.: 1973, *Nucleotide Metabolism: an introduction*, Academic Press, New York

Higgins,D.G. and Sharp,P.M. (1989). *CABIOS* **5**, 151-153.

Holm, L., and Sander, C.: 1996, *Proteins* (*en prensa*)

Horowitz, N. H.: 1945, *Proc. Natl. Acad. Sci. USA* **31**, 153-157

Horowitz, N.H.: 1965, In Bryson, V., and Vogel, H.J. (eds*), Evolving Genes and Proteins*, Academic Press, New York, 15-23

Iwabe, N., Kuma, K., Hasegawa., M.,Osawa, S., and Miyata, T.:1989, *Proc. Natl. Acad. Sci. USA*, **86**, 9355-9359

Jensen, R.A.: 1976, *Annu. Rev. Microbiol.* **30**, 409-425

Keefe, A.D., Lazcano, A., and Miller, S.L.: 1995, *Origins Life Evol. Biosphere.* **25**: 99-110

Koonin, E.V., Tatusov, R.L., and Rudd, K.E.: 1995, *Proc. Natl. Acad. Sci. USA* **92**, 11921-11925

Labedan, B. and Riley, M.: 1995, *Mol. Biol. Evol.* **12**, 980-987

Lazcano A.: 1993, *Trends in Microbial Ecology*, 559-562

Lazcano, A. and Miller, S.L.:1994, *J. Mol. Evol.* **39**,546-554

Lazcano, A. and Miller, S.L.: 1996, *Cell* **85**, 793-798

Lazcano ,A., Fox, G.E., and Oró, J.: 1992, *The Evolution of Metabolic Function*, CRS Press, Boca Raton.

Lazcano, A., Díaz-Villagomez, E., Mills, T., and Oró, J.:1995, *Adv.Space Res.* **15**,345-356

Munch-Petersen, A (ed): 1983, *Metabolism of Nucleotides, Nucleosides, and Nucleobases in Microorganisms*, Academic Press, London

Mushegian, A.R. and Koonin, E.V.: 1994, *Protein Science* **3**,1081-1088

Mushegian, A.R. and Koonin, E. V.: 1996a, *Proc. Natl. Acad. Sci. USA* **93**, 10268-10273

Mushegian, A.R. and Koonin, E.V.: 1996b, *TIGS* **12**, 289-290

Oparin, A.I.: 1938, *The Origin of Life*, Macmillan, New York

Pennissi, E.:1996, *Science* **272**,1098-1099

Riera, j., Robb F.T., Weiss, R., and Fontecave, M.:1997 *Proc. Natl. Acad. Sci. USA (en prensa)*

Silver, W.S. and Postgate, J.R.: 1973, *J. Theoret. Biol.* **40**, 1-10

St. Jean, A. and Charlebois, R. L.: 1996, *J.Bacteriol.* **178**, 3860-3868

Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, M.S., Borodovsky, M., Rudd, K.E., and Koonin, E.V.: 1996, *Current Biol.* **6**, 279-291

Tham, T.N., Ferris, S., Kovacic, R., Montagnier, L., and Blanchard, A.: 1993, *J. Bacteriol.* **175**, 5281-5285

Waley, S.G.: 1969, *Comp. Biochem. Physiol.* **30**, 1-7

Weber, G., Nakamura, H., Natsumeda, Y., Szekeres, T., and Nagai, M.: 1992, *Adv. Enzyme Regul.* **32**, 57-69

Woese C.R.: 1983, In Bendall, D. S., K. (ed.), *Evolution from Molecules to Man*, Cambridge University Press, Cambridge, 209-233

Woese C.R.:1987, *Microbiol.Rev.* **51**, 221-271

Woese, C.R. and Fox, G.E.: 1977, *J.Mol.Evol.* **10**: 1-6

Ycas, M.: 1974, *J. Theoret. Biol.* **44**, 145-160

Zubay, G.: 1993, *Chemtracts - Biochem. Mol. Biol.* **4**, 317-323