Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Evolution without speciation but with selection: LUCA, the Last Universal Common Ancestor in Gilbert's RNA world

Hugo Hoenigsberg

Institute of Evolutionary Genetics and Molecular Biology,
Universidad Manuela Beltran, Bogotá D.C., Colombia
Corresponding author: H. Hoenigsberg
E-mail: hoenisbe@umb.edu.co

Genet. Mol. Res. 2 (4): 366-375 (2003)
Received August 6, 2003
Accepted October 9, 2003
Published December 8, 2003

TO THE MEMORY OF H.J. MULLER, TH. DOBZHANSKY, C. BARIGOZZI AND R. CEPPELLINI, WHO IN DIFFERENT WAYS SUGGESTED SOME ASPECTS DEVELOPED IN THIS THEORY

ABSTRACT. This is not an attempt to analyze the Last Universal Common Ancestor (LUCA) to understand the origin of living systems. We do not know what came before Gilberts' RNA world. Our analysis starts with the RNA world and with genes (biological replicators *alla* Dawkings) made up of RNA proteins with enzymatic catalytic functions within units that are not yet modern cells. We offer a scenario where cellular entities are very simple and without individuality; they are only simple primary units of selection (the first level of selection) in which replicators compete in the most Darwinian manner, totally deprived of cooperation and interactions among genes. The information processing system of this RNA world is inaccurate and inefficient when compared to that found in organisms that came later. Among the "genes" and the entities that harbor them, high mutation rate was the most prevalent source of variability and the only inheritance was through lateral gene transfer of mobile elements. There were no chromosomes or any other genomic organization. As millions of years accumulated, complex and organized biological

structures and processes evolved thanks to the variability mustered up mostly by lateral gene transfers and mutations. With micro- and minisatellites, lateral gene transfers became indispensable devices of selection to mold variability. Competition and Darwinian selection gave way to a new transition in evolution, one I consider ineluctable, in which cooperation among interactive genes prevailed for the sake of higher fitness. Compartmentalization constituted a major transition in evolution that spurted new types of genome organization. Minichromosomes is one of these; cellular membranes and cytoplasmic structures completed the picture of the primitive cell. However, the much talked about phylogenetic tree does not exit in that ancient LUCA. The tree has no organism at its base; only clusters of genes evoke a fragile beginning for the increasingly complex cell types that were to emerge later.

Key words: Chemical replicators, Evolving genes, Gene clusters, Molecular Cooperation, Molecular interactions, Aggregates, Before phylogenetic tree

INTRODUCTION

To be subjects of evolution, and through it, of the different levels of selection, entities have to meet the following criteria: 1) They have to pass their characteristics on to future generations (heredity); 2) By whatever means, they have to multiply their own kind (multiplication), and 3) Their heredity must have variants (variability). For Darwinian evolution to be effective, entities have to produce hereditary differences in fecundity and/or survival. Differentials in fertility, fecundity, longevity, infant mortality, sexual and sensorial discrimination, are unfailing elements in the various kinds of selective mechanisms. These populational, physiological and molecular characteristics are part of the modern evolutionary process of today's faunas. However, these biological ingredients were not always there to further changes in gene frequencies.

The initiation of life did not enjoy these three blessings. When chemical entities were the principal reactants you could have had multiplication without evolution because there was no variation among the products. Szathmáry (1999) argues that to have an evolutionary process going you have to have "unlimited heredity" of template copying of Dawkins' replicators. The first steps of evolution were not accompanied by natural selection because there was not unlimited heredity. Szathmáry uses the "formose" reaction (the autocatalytic synthesis, catalyzed by an enzyme from outside, of formaldehyde to glycolaldehyde, first discovered by Butlerov in Russia one hundred years ago) in which the core reaction is the spontaneous generation of glycolaldehyde. At the dawn of life small organic molecules could, in each cycle, multiply their kind without heredity.

Generally, biological growth follows Malthusian dynamics of exponential growth, without biotic or abiotic limitations. The theoretical analysis of selection that intends to demonstrate its effectiveness assumes that growth is exponential. Nevertheless, in the chemical construct of test tube replicators by Kiedrowski (referred to by Szathmáry, 1999) template and copy follow a slower than exponential rate of growth because they are bound together by hydrogen bonds;

meaning that only when templates are free can they initiate a new cycle of replication. For example, when RNA molecules replicate in a test tube replicase efficiently separates the template and the copy so that both can enter into the new cycle of replication. This later growth is exponential, but the initial growth is not, in view of the fact that it is self-inhibited. Szathmáry contends that one of the important escalations of prebiotic evolution must have been replicators with exponential growth. But before replicators of the RNA world can have template exponential growth many molecular refinements have to be assumed and accepted.

With Kiedrowski's growth limitation of templates we are still left with the big question of how does natural selection and heredity enter into the picture of primitive genome organization? We address this ambiguous situation next, conscious of the time element involved. Without the extraordinary velocity that selection and heredity impel and stamp on the evolutionary process, we have to assume several million years. Thus, Kiedrowski's growth limitation for the first step implies a slow molecular evolution, indispensable to evolve free template replication cycles of oligonucleotides in the RNA world.

The next step was for evolution to evolve, using heredity and selection, was nucleotide replication with enzymatic properties (Gilbert's RNA world; Gilbert, 1986), although unstable at first, to protect local informational chemistry from relentless molecular-parasitic invasions. To answer the challenge of nearby molecular invasions, genes organized themselves into larger genomes as minichromosomes, and "jumped" to a new level of cooperative selection by creating cell membrane and nuclear compartments. The gist was to trade off higher level fitness, for more competitive, small, selfish replicators of the previous level, similar to todays' primers (with a few nucleotides). Minichromosomal organization acquired considerable fitness reinforcement in the new higher level of selection, with the concomitant evolution of epistatic interactions.

LUCA IN GILBERT'S RNA WORLD

Nothing concrete can be said apropos LUCA's physical appearance, but the common ancestor can be perceived as a molecular entity invested with information qualities. For instance it can be conceived as a member of a phylogenetic line of descent without an organismal corporeal existence, without genealogy, similar to single genes of the RNA world, loosely united in a network with the evolvable gene clustering that in time encoded chaperone proteins, known as enzymes, to develop into a growing functioning metabolism. Epigrammatically, this had to be a simple genetic entity, without a real intermediary metabolism to accompany its beginning. In this view, I see LUCA as deprived of complex protein capabilities as the result of a deficient information processing apparatus. For example, complex proteins, like primases, helicases, DNA polymerases and other familiar enzymes in membrane and nuclear compartments, are out of the question. The replication process that we see in today's eukaryotes could not have existed. Only chaperone proteins can be expected from this common and communal ancestor. It was nothing but a genetic network of RNA genetic units in total Darwinian "war" among themselves; this was the first theater of selfish genes, with little or no-intermediate metabolism. LUCA "lived" without any of the cell compartments, totally at the mercy of lateral "abuses" from nearby oligonucleotides, molecular parasites that endangered its "incipient library". Among regular genes, cancer-like selfish genes were found in that strange milieu, but without telomerases doing what they do in reconstructing the tip of the extended chromosomes. Minichromosomes came later as structures disconnected from mitosis. To emerge from that first level of natural selection, a

new major transition in evolution had to occur (Maynard Smith and Szathmáry, 1995). Evolutionary transitions in fitness and individuality (Michod, 1999) have always occurred when cellular and ecological conditions have pressured organisms' genomes. Therefore, the presence of telomerases should be left to a higher evolutionary transition where some individuality was acquired. After all, genomes had to be protected and minichromosomes had to be kept from being sticky.

As evolution proceeded (during millions of years) selection was felt in various departments; for instance, although lateral genes were transferred from one group to another, as the translation mechanics became more efficient (less mistakes in identifying codons) and more complex proteins were made, gene clustering became vital and natural selection certainly acted to aggregate previously separated loci (Demerec and Hartman, 1959). Encoded enzymes needed to work on similar substrates, which allowed "the assembly line of genes" to be present at the very beginning of a proper metabolism (Pontecorvo, 1950).

We assume that these genetic entities were made of ribosome RNA with information and catalytic capabilities. The properties of RNA proteins, plus the lateral transfers, which in those days must have been considerable, selection in favor of gene clustering and high mutation rate, did not permit isolation among networks of genes. Conditions for selection to act on local ecological conditions, and later to profit from mechanisms of allopatric speciation were not on hand. Lateral gene transfers homogenized genomes and gene clusters. We are proposing an ancient stage, in the Archean, before 3.5 billion years ago, when the proto-informational world was experimenting with cellular biology and nucleic acid-protein translation. Evolution was just wrestling with single discrete units of nucleic acids and elemental proteins, but practically no intermediate metabolism at the root of the phylogenetic tree of life. The network of related genes in a community without a common evolutionary design or cooperation, without orientation, and without individuality, so to speak, survived and evolved as an inaccurate biological unit. Considering the 3.5 billion years of evolutionary distance from today, we cannot avoid standing in awe thinking of the fragility of this protocellular animal as a crippled riboorganism that emerged from ribosomal RNA but which, on the other hand, spawned the three most remarkable taxonomic domains of life, the Archea, the Bacteria and the Eucarya.

BEFORE ARCHEAL GENOMICS

Our purpose in this essay is not to recapitulate what has been found in test-tube evolution with dual-purpose genetic enzymes made with either RNA or DNA (as was reported by Yale University's Ronald Breaker and Adam Roth in 1998). The undoubted biochemical skill that went into making a tailor-made enzyme as the first nucleic acid enzyme that uses an amino acid to trigger chemical activity is in itself something to reckon with. There are no naturally occurring DNA enzymes to date. But remembering that approximately two decades ago RNA enzymes were discovered as ribozymes by Sydney Altman and Thomas Cech (Nobel Price in 1989 in Chemistry), we should not be surprised if DNA enzymes are found outside labs.

Anyway, as evidence is mounting that it was an RNA world at the dawn of life when the Earth cooled down, it is not amiss to theorize what could have happened at the origin of the archaeal genomics phylogenetic root that initiated that first stage in the evolution of life. Recent data on archaeal genomics (Doolittle and Logsdon, 1998) interpret a third complete archaeal genome sequence with eukaryote-like genes for replication, transcription and translation; no doubt this is a revolution in microbial evolution since it makes them members of mixed origins.

The amazing story of the "complete genome sequence of the hyperthermophilic sulfate-reducing archeon *Archaeoglobus fulgidis*" (Klenk et al., 1997), has forced every student of evolution to come to grips with the great discovery of the archaea in the course of constructing a universal evolutionary tree from ribosomal RNA sequence information (Woese and Fox, 1977a,b). Thus, archaebacteria and its progenote in the universal root have a special evolutionary position and constitute a phylogenetically coherent group (Doolittle and Brown, 1994).

Others have proposed a sequential Darwinian model for the evolution of life, taking as the starting point the late moments of the RNA world, going on to the origins of eukaryotes and prokaryotes (Jeffares et al., 1997; Poole et al., 1998). The apparent simplicity of prokaryotes and a great trust in comparative phylogenetic sequence data used in tree-building methods, have led many to consider them as predating eukaryotes, notwithstanding considerable debate (Doolittle, 1995; Baldauf et al., 1996). Archaea and eubacteria are fundamentally similar in genome organization (Baumann et al., 1995) and we do not quarrel with this well-established position simply because we place ourselves even before taxonomy. It is possible that the strong argument that uses Darwinian r selection as the instrument of change based on relics from the RNA world, which arguably emphasizes that the prokaryotic genome organization is derived (Poole et al., 1998) and that they passed through a thermophilic stage, as suggested by Forterre (1995) and others, have to be reckoned with. Therefore, we could have today's prokaryotes as a group that acquired its smaller compact streamlined advanced genome, and the circular plasmid-like molecule, after the thermophilic stage (Poole et al., 1998).

We address another aspect of the original LUCA: that one stage in the evolution of the universal common ancestor in which only selfish genes were present, at first without an organismal or genealogical identity.

THE PHENOTYPE OF THE UNIVERSAL COMMON ANCESTOR

What we have in the way of phenotype is a network of clones of ribosome genes (rRNA-like) that had to "live" with its unstable nature, before a more stable DNA version was fabricated with a mechanism for making RNA copies assisted by natural selection. At the dawn of life the cripple protein-synthesizing ribosome and its relatively high fidelity RNA polymerase that joined triplets of oligonucleotides initiated its evolutionary career in enzymatic activity as a catalytic apparatus from RNA to nonspecific protein. Fossil RNA in the "habitat" of the RNA world were bits of RNA, ancestors of todays RNA introns splicing themselves out of genes, remnants of other bits that in us are guiding scouts of RNA, vaulted RNA, small nuclear RNA, small nucleolar RNA, and self splicing introns. Today's bacteria have none of these; maybe they dropped them as they evolved to their present day highly advanced state. Having specialized in scavenging and parasitic niches with an r strategy, good at Darwinian competitions with high-speed reproduction, bacteria evolved a new streamlined machinery, with none of the fossil RNA that eukaryotes retained from our ancient universal ancestor (see Poole et al., 1998).

No one can claim complete clarity at the root of the phylogenetic tree. The much reputed paralogous gene group method used frequently (Gogarten et al., 1989; Iwabe et al., 1989 and others) suggests that the ancestor is a specific organism. This is a big order in our view, because it means that it can be a node on a tree. Can Archaea be related to the eukaryotes? It is unreasonable to consider the universal common ancestor a prokaryote, meaning both Archaea and Bacteria!

The sequences in the three domains have turned out to be anything but straightforward in an already difficult scenario (Gogarten et al., 1989; Iwabe et al., 1989; Gupta and Singh, 1994; Lazcano and Miller, 1994; Baldauf et al.,1996). Other workers have found these results confusing (Shiba et al., 1997; Brown and Doolittle, 1997; Woese, 1998). Whatever the confusion, one thing should orient a parsimonious discussion for the sake of Occam's razor, that LUCA has to be a simple ancestor, limited in its metabolism and in its energy source. Before the three domains diverged, the cluster of genes in the initial network with just chaperone proteins and no cellular compartmentalization had to increase their informational file mostly through lateral transfers and mobile elements. The ribosome gene network can be enriched all the time, even before cellular functions.

I suppose that with target primed reverse transcription (TPRT), L1s retrotransposons can also contribute to enrich genomic RNA, even in the universal ancestor, but at a stage beyond our initial cluster of ribonucleic genes. Thus, diversification of their lineages can ensue without direct transfer within and among taxa.

We cannot avoid Demerec and Hartman's theory (1959), which says that regardless of how gene clusters originated, natural selection worked (even from LUCA on) to aggregate previously separated loci. After all, distantly related organisms that present homologous proteins reveal the relationship among the genes that coded them (Zuckerkandl and Pauling, 1962).

PREVIOUS TO TRANSLATION

Assuming an early and primitive beginning, ulterior change requires instruments such as natural selection, mutation, sample variance and drift, and whatever makes the accumulation of diversity among lines possible. On the other hand, one of the first elements of change and of efficient integrations is molecular reproduction and homologous recombinations. Are we, as theoretical evolutionists, authorized to start with molecular evolvability that optimized physiological relativity and cooperation among genes? Molecular reproduction and homologous recombination to selectively multiply dispersal of advantageous alleles to high frequencies among genetically related clusters was the first biological arrangement to step out of the incorporeal, somewhat undefiled and definitively intractable stage of the first level of selection. There was a trade off between the first Darwinian level of selection (selfish genes prevailing) and the higher fitness of efficient molecular integrations. It was also the first attempt to ward off against selfishness and to enter into cooperative arrangements among genes. We can imagine this as the first jump into the unknown, meaning into a new level of selection, into "discipline", as integrated clusters of genes alla Pontecorvo (1950). To become a good assembly line the second level of selection gradually joined closely together what had been a loosely and often a distant ensemble of gene clusters.

The second contrivance was lateral gene transfer, which introduces novel information in the RNA world, even from distant unrelated protolineages via illegitimate recombination. We will never know where these lineages *in status nascendi* could have taken the first group of organisms. We can only offer an explanation for the emergence of gene exchange by this route, out of logic, because it serves to distribute genes throughout distantly related groups (later, future clades) and therefore may confer complex abilities and higher fitness. Closely related protolineages accumulated sameness. These two mechanisms of gene exchange played complementary roles in the diversification of early life. Although no one can speak of "species" a cogent process of speciation was on its way.

MICROSATELLITES AS PROVIDERS OF CREATIVITY

The early function of SSR, STR or VNTR was not junk, even though it maybe today in some eukaryotic nuclei. Microsatellites had an important function in the RNA composition of LUCA, when high mutation rate and selection of protoorganism genomes endeavored to cope with insidious and constant environmental changes of the Archean period (about 4000 to 3800 million years ago). Shifts in osmotic pressure, temperature and energy changes, water tides, etc., endangered incipient membranes and independent genes with recently acquired alleles or mobile elements in microchromosomes. Thus, microsatellite alleles in coding regions could have been the early mechanism to get quick adaptive responses: STR or SSR could aid genomes in one environment and another allele, of either one, in another. Therefore, having variation in tandem repeats ensured the survival of the population of gene clusters before genetic drift and selection erased much needed variability in the local population.

As a mechanism to compensate for loss of genetic variability due to drift and negative selection, microsatellite variation within populations not only serves to ensure survival in varying environments, but is also a means to produce some versatility in protein function and gene expression. Microsatellites are systematically found near coding regions; their allelic variation has been shown to be associated with quantitative variation in protein function and gene activity.

Today, in humans, 90% of known microsatellites are found in noncoding regions of the genome. When found in human coding regions, they are associated with disease. Nevertheless, it is possible that in early days microsatellites were selectively promoted to cause rapid adaptive responses, as we have suggested.

To go through such variable environments in the Archean, favored coefficients of selection acted primarily on evolvable characteristics, those that produced changes that were not detrimental *prima facie* to cellular development. The features that increased the chances of becoming individuals with organismic identity, beyond simple genes, were those, such as a nuclear membrane, complex cell membrane with different types of molecular signals, several pumps to equilibrate osmotic pressure and several internal organelles to carry on transcription and translation, sufficient for the emergent-incipient RNA world, but which later on became indispensable to cope with more sophisticated genomes and complex protein developments. These features did not emerge all at once. Before cellular compartmentalization the RNA translation mechanism had to evolve to one able to deal with codons without making mistakes. Thus, more complex proteins came forth in the next transition in evolution.

In the days of the early ancestor most of these functional compartments were effectively isolated thanks to lipid membranes, across which most materials moved selectively. We are aware that only a gradualist natural selection could have operated then, as today in some instances. These evolvable changes in general were not abrupt; however, we are investing the early ancestor with the capacity not to abrogate some apparently dangerous mobile elements that enter via lateral transfers, and to assume oligonucleotides that could have been part of transitory semiloads (semilethals). The mechanism to retain non-adaptive loads was for the genome to accept those that did not produce rebuff. Most of the genes in the network were clones invested with a considerable degree of consanguinity. There was a great deal of compatibility necessary for gene complementation and interactions. Gene modifiers of various kinds that police detrimental repeats were the only mechanisms within the genome for the molecular rejection of genes that had no capacity for interactive behavior. At the beginning there was no mitosis.

Microsatellites and recombination were early devices to ward off sameness.

Kimura's molecular neutrality produced genetic and chemical optima that constrained selection, because it (selection) can harm the chemical machinery and through it become detrimental to cell biology.

If the original scenario was a group of clones in the midst of lateral gene transfers and mobile elements carrying other elements that in time developed some degree of local variability through mutations and ecological diversification, we have the way paved for co-adaptation of gene complexes to facilitate the coordinated expression and regulation indispensable for the initiation of metabolism. Thus, selection paved the way toward the evolution of gene complexes by favoring primitive operon structure in minichromosomes.

THE EVOLUTION OF THE GENETIC CODE

The greatest challenge to the origin of life is the evolution of translation. All extant life forms use protein enzymes, and their fabrication results from translating an RNA message. Invariably, translation takes place in a highly complex RNA-protein apparatus, the ribosome, using tRNAs that are specifically adapters to an amino acid. All present day organisms (as far as we know) use the same set of 20 amino acids. In all cases, tRNAs function as adapters that permit the transfer of an amino acid to the growing chain whenever they find that the consecutive nucleotides (triplets) that form the codon of the mRNA match the three anticodon nucleotides of the tRNA. Aminoacyl-tRNA synthesis typically is made by the 20 aminoacyl-tRNA synthetases, each one specific for an amino acid (there are some exceptions).

Although the code itself is more or less invariant, there are certain features, which frequently appear in the experimental evidence, that force one to address the problem of the action of natural selection at its very origin (see Crick, 1968; Orgel, 1968): 1) The 20 amino acids are not distributed at random among the 64 triplets. Following Crick (1968), XYU and XYC always code the same amino acid. 2) XYA and XYG often code the same amino acid. 3) The rare amino acids, methionine and tryptophan, have only one codon each.

See Table 1 in Crick's 1968 paper showing the "best allocations" of the 64 codons. There are a great many more experimental papers that show the non-randomness of the amino acids "picked" for the 64 codons available. What follows is that very early in LUCA's existence natural, selection probably acted: 1) on the strength of the effectiveness of the codon-anticodon match, 2) on the relationship of the codon-amino acid stereo-chemical necessity (see Woese, 1967), and 3) because the match of those specific amino acids and their codons were the viable ones.

Maybe these three reasons account for the fact that the code does not change today and that the translating apparatus is universal (more or less). However, these three reasons also serve to tell us that if the code determines highly evolved protein molecules today, and that any change to these would be highly deleterious to their carriers, then ancient simple proteins did not require precise matching simply because natural selection did not have to be so demanding. Our theory for LUCA is that at its origin, ribogenetic evolution did not begin with the same strength; it was slow, because selection was not as demanding. Most proteins were simple, not complex; their cellular implications were not meant to be anything close to an immaculate involvement. There were times when codes were not fully used, thus they were not selectively accommodating, i.e., the coefficients of selection were low, the very mediocre ensemble of amino acid

anticodon fitting was sufficient for global survival of homogeneous genomes. Much of the amino acid anticodon fitting was really adventitious.

FINAL REMARKS

A recent review about the RNA world (Yarus, 1999) serves well those who want to make up their minds about the origin of the genetic code. However, we realize that other simpler molecules could do the trick in prebiotic existence (Knight and Landweber, 2000; Knight et al., 2001a,b). PNAs have been offered rather than RNA as the first genetic molecule (Nelson et al., 2000). What we have explored in our review initiates with a specific model of a protocell, just after tRNA-based peptide synthesis results as an inevitable mechanistic model of Gilbert's RNA world. The ribosome protein-enzyme catalysis is used for replication. Our LUCA starts before others (Penny and Poole, 1999). Our LUCA is not prebiotic, but it is not as advanced as those that are looking for a node in a phylogenetic tree.

Therefore, the main characteristics of LUCA's still fragile translation mechanism are: 1) The RNA genome. 2) the RNA ribosome. There is considerable *in vitro* evidence (Nitta et al., 1998) and by analysis of atomic structure (Ramakrishnan and Moore, 2001) that the ribosome is essentially a ribozyme. 3) That tRNA has always been an adaptor in the translation apparatus. 4) The ribozyme aminoacyl synthetases have been demonstrated by means of *in vitro* evolution (Lee et al., 2000); moreover, there seems to be evidence that tRNAs predated their synthetases (Ribas de Pouplana et al., 1998). 5) Protein replicase activity (Johnston et al., 2001) and ribozymes with ligase-based replication functions (McGinness and Joyce, 2002) were recently obtained *in vitro*!

ACKNOWLEDGMENTS

The author thanks two referees for opinions and corrections that made this hypothesis more acceptable.

REFERENCES

Baldauf, S.L., Palmer, J.D. and **Doolitlle, W.F.** (1996). The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl. Acad. Sci. USA 93*: 7749-7754.

Baumann, P., Qureshi, S.A. and **Jackson, S.P.** (1995). Transcription: new insights from studies on archaea. *Trends Genet. 11*: 279-283.

Brown, J.R. and **Doolittle, W.F.** (1997). Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev. 61*: 456-502.

Crick, F.H.C. (1968). The origin of the genetic code. *J. Mol. Biol. 38*: 367-379.

Demerec, M. and **Hartman, Z.E.** (1959). Complex loci in microorganisms. *Annu. Rev. Microbiol. 13*: 377-406.

Doolitlle, R.F. (1995). The origins and evolution of eukaryotic proteins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci. 349*: 235-240.

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

Doolitlle, W.F. and **Logsdon, J.M.** (1998). Do archaea have a mixed heritage? *Curr. Biol. 8*: R209-R211.

Forterre, P. (1995). Thermoregulation, a hypothesis for the origin of prokaryotes. *C.R. Acad. Sci. Paris Life Sci. 318*: 415-422.

Gilbert, W. (1986). The RNA world. *Nature 319*: 618.

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J.,

Date, T., Oshima, T., Konishi, J., Denda, K. and **Yoshida, M.** (1989). Evolution of the vacuolar H^+ ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA 86*: 6661-6665.

Gupta, R.S. and **Singh, B.** (1994). Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic nucleus. *Curr. Biol. 4*: 1104-1114.

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

Jeffares, D.C., Poole, A.M. and **Penny, D.** (1997). Relics from the RNA world. *J. Mol. Evol. 45*: 18-36.

Johnston, W.K., Unrau, P.J., Lawrence, M.J., Glasner, M.E. and **Bartel, D.P.** (2001). RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science 292*: 1319-1325.

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.R., Peterson, J.D. et al. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature 390*: 364-370.

Knight, R.D. and **Landweber, L.F.** (2000). The early evolution of the Genetic Code. *Cell 101*: 569-572.

Knight, R.D., Freeland, S.J. and **Landweber, L.F.** (2001a). Rewiring the Keyboard: evolvability of the genetic code. *Nat. Rev. Genet. 2*: 49-58.

Knight, R.D., Freeland, S.J. and **Landweber, L.F.** (2001b). A simple model based on mutation and selection explains trends in codon and amino acid usage and GC composition within and across genomes. *Genome Biol. 2*: 1-13.

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

Lee, N., Bessho, Y., Wei, K., Szostak, J.W. and **Suga, H.** (2000). Ribozyme - catalyzed tRNA aminoacylation. *Nat. Struct. Biol. 7*: 28-33.

Maynard Smith, J. and **Szathmáry, E.** (1995). *The Major Transitions in Evolution*. W.H. Freeman, San Francisco, CA, USA.

McGinness, K.E. and **Joyce, G.F.** (2002). RNA-catalyzed RNA ligation on an external RNA template. *Chem. Biol. 9*: 297-307.

Michod, R.E. (1999). *Darwinian Dynamics*. Princeton University Press, Princeton, NJ, USA.

Nelson, K.E., Levy, M. and **Miller, S.L.** (2000). Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. USA 97*: 3868-3871.

Nitta, L., Kamada, Y., Noda, H., Ueda, T. and **Watanabe, K.** (1998). Reconstitution of peptide bond formation with *Escherichia coli* 23S ribosomal RNA domains. *Science 281*: 666-669.

Orgel, L.E. (1968). Evolution of the genetic apparatus. *J. Mol. Biol. 38*: 381-393.

Penny, D. and **Poole, A.** (1999). The nature of the last common ancestor. *Curr. Opin. Genet. Devel. 9*: 672-699.

Pontecorvo, G. (1950). New fields in biochemical genetics of microorganisms. *Biochem. Soc. Symp. 4*: 40-50.

Poole, A.M., Jeffares, D.C. and **Penny, D.** (1998). The path from the RNA world. *J. Mol. Evol. 46*: 1-17.

Ramakrishnan, V. and **Moore, P.B.** (2001). Atomic structures at last: the ribosome in 2000. *Curr. Opin. Struct. Biol. 11*: 144-154.

Ribas de Pouplana, L., Turner, R.J., Steer, B.A. and **Schimmel, P.** (1998). Genetic code origins: tRNAs older than their synthetases? *Proc. Natl. Acad. Sci. USA 95*: 11295-11300.

Shiba, K., Motegi, H. and **Schimmel, P.** (1997). Maintaining genetic code through adaptations of tRNA synthetases to taxonomic domais. *Trends Biochem. Sci. 22*: 453-457.

Szathmáry, E. (1999). The first replicators. In: *The Levels of Selection* (Laurent, K., ed.). Princeton University Press, Princeton, NJ, USA, pp. 31-52.

Woese, C.R. (1967). *The Genetic Code*. Harper & Row, New York, NY, USA.

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

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

Woese, C.R. and **Fox, G.E.** (1977b). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA 51*: 221-271.

Yarus, M. (1999). Boundaries for an RNA world. *Curr. Opin. Chem. Biol. 3*: 260-267.

Zuckerkandl, E. and **Pauling, L.** (1962). Molecular disease, evolution and genetic heterogeneity. In: *Horizons in Biochemistry* (Kasha, M. and Pullman, B., eds.). Academic Press, New York, NY, USA, pp. 189-225.