[Frontiers in Bioscience 13, 5605-5613, May 1, 2008]

The genomics of LUCA

Wai-Kin Mat, Hong Xue, Jeffrey Tze-Fei Wong

Department of Biochemistry and Applied Genomics Center, Hong Kong University of Science & Technology, Clear Water Bay, Hong Kong, China

TABLE OF CONTENTS

1. ABSTRACT

To understand the nature and evolution of LUCA, or Last Universal Common Ancestor, the minimum genome of LUCA has been identified based on the genes common to the eight primitive *Euryarchaea* and *Crenarchaea* species *Methanopyrus kandleri*, *Methanothermobacter thermautotrophicum*, *Methanococcus jannaschii*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeropyrum pernix* and *Pyrobaculum aerophilum*, together with the methanogenesis genes of the primitive methanogens. The 424 protein encoding genes in the minimum LUCA genome exceed significantly the 150-340 genes estimated to be present in a minimal proteome compatible with life. Thus LUCA was not a minimal organism but the first modern organism equipped with a DNA genome and the universal genetic code. The hyperthermophilic, *Methanopyrus*-proximal LUCA is consistent with a Hot Cross Origin of life which proposes that early heterotrophic life forms in the cooler temperature zones invented methanogenesis and a DNA genome upon their adaptation to the hydrothermal vents, where life flourished massively on lithoautotrophy supported by carbon dioxide and hydrogen, thereby leading to the rise of LUCA.

2. ROOT OF LIFE

Locating the root of life is a prerequisite to the analysis of early evolutionary events. Not knowing where the root is located, the nature of the root cannot be determined, the evolutionary relationships between the three domains of life cannot be analyzed, and the earliest branchings from the root cannot be traced. Accordingly, there has been an intense search for the root in recent years. The main approach employed in the search has relied on the mutual rootings of paralogous proteins generated by gene duplications that took place prior to the appearance of the root organism, viz. the Last Universal Common Ancestor (LUCA). However, on account of artifacts generated by horizontal gene transfers, long branch attraction, mutational saturation and erratic species, this method has given unreliable results, thereby giving rise to deep pessimism in the field (1-3).

To avoid the difficulties of paralogous proteins, there are three other sequence information-rich biopolymers one may turn to, which are DNA, rRNAs and tRNAs. Of these DNA and rRNA in a cell are devoid of paralogs, and therefore cannot be employed for rooting. Single tRNA sequences, owing to their short lengths, also

were not found to be useful. This situation, however, has been fundamentally altered by the complete sequencing of genomes, which yields not just single tRNA sequences, but a complete complement of tRNAs in the genome with a typical minimum of 32 tRNAs 75 bases long, or a total of 2,400 bases. In addition, tRNA sequences turn out to evolve more slowly than proteins, and be less subject to variations caused by lifestyle adaptations, such as a choice between the parasitic and free-living modes. There is also a paucity of horizontal gene transfers of tRNA exons. These attributes in combination have rendered the tRNAs uniquely valuable phylogenetic probes especially over long time spans, as required for the search for LUCA. As a result, polyphasic evidence based on tRNA sequences and supplemented by a wide range of additional lines of evidence has located LUCA close to the hyperthermophilic methanogen *Methanopyrus kandleri* (Mka) (4-8).

3. THE LUCA GENETIC CODE

The utilization of basically the same universal genetic code by all extant organisms indicates that they all descended from a LUCA that possessed this code. This requires that all of LUCA's contemporary lineages that employed alternate codes came to be thoroughly eliminated in time. Biological species come and go, and mass extinctions are not unknown, but the elimination of all contemporary species by a single lineage is unknown in biological records aside from the ascendance of LUCA, which was therefore a unique event in the history of the living world.

In the early stages of prebiotic evolution it might be difficult to resolve the biotic populations into distinct species. However, the establishment of the universal genetic code suggests that LUCA was a clearly delineated species at least with respect to its code. *Thus LUCA may be defined as the first organism endowed with the universal genetic code of present day organisms*.

Taking the genetic code of *Methanopyrus* as the closest description of the LUCA code, the LUCA code distributed codon assignments to 20 standard amino acids, and lacked GlnRS, AsnRS and CysRS (aminocyl-tRNA synthetases for Gln, Asn and Cys), such that its Gln-tRNA had to be synthesized through amidation of Glu-tRNA, Asn-tRNA through amidation of Asp-tRNA, and Cys-tRNA through transformation of phosphoSer-tRNA (8). All of the thirteen standard 1-amino acid or 2-amino acid codon boxes (viz. UUN, UCN, CUN, CCN, CAN, CGN, ACN, AAN, AGN, GUN, GCN, GAN and GGN) were most likely read by two tRNA species per box bearing the GNN and UNN anticodons (6).

4. LUCA WAS NOT A MINIMAL ORGANISM

Gene contents vary extensively among organisms, indicating that they closely track the adaptation of various species to their ecological niches. Accordingly, even though Mka is phylogenetically proximal to LUCA, significant differences are expected between the Mka and the LUCA genomes, and it becomes necessary to identify the composition of the LUCA genome based on the genes common to the most primitive species. Earlier, the finding that the six ancient archaeal species Mka, *Methanothermobacter thermautotrophicum* (Mth), *Methanococcus jannaschii* (Mja), *Pyrococcus abyssi* (Pab), *Pyrococcus furiosus* (Pfu) and *Pyrococcus horikoshii* (Pho) are lacking in cytochrome genes has led to the suggestion that LUCA did not possess any cytochrome genes (8). The rationale is straightforward. These *ancient six* species fall into two opposing metabolic groups: Mka, Mth and Mja feed on hydrogen and carbon dioxide, whereas Pab, Pfu and Pho produce hydrogen and carbon dioxide as metabolic end products. Therefore the cytochrome deficiency common to the two groups is ascribable to shared primitivity rather than metabolic similarity. On this basis, the genes that are common to these *ancient six* may be regarded as primitive genes that were likely to be constituent genes of the LUCA genome.

Among the *ancient six*, because the Mka-Mth-Mja group and the Pab-Pfu-Pho group have dissimilar pathways of energy metabolism, the genes common to the two groups would lack the specific energy metabolism genes from either group. Because LUCA could not survive without any specific energy metabolism genes, such genes need to be added to the LUCA gene set. Since the primitive methanogens are closest to LUCA (4-8), the methanogenesis genes that are common to Mka, Mth and Mja are included in the plausible proteome of 561 LUCA COGs (Clusters of Orthologous Groups) (9) in Appendix 1. Since some LUCA genes could have been deleted in any one of the *ancient six*, this set of 561 genes gives a low end estimate for the LUCA proteome. For example, Mka and most likely LUCA contain the SelD (Mk1369) and SelA (Mk0620) genes encoding the pretran synthesis of Cys-tRNA (8), but this pathway has been abandoned in some of the *ancient six*. Accordingly these LUCA genes do not appear in Appendix 1.

LUCA is located between the *Euryarchaea* and *Crenarchaea* on the universal tRNA phylogenetic tree. Since the *ancient six* are all primitive euryarchaeons, the COGs common to the *ancient six* might include genes that were originally absent from LUCA but were added to its earliest euryarchaeal offsprings. To counter this possibility, Appendix 1 also shows the smaller proteome of 424 COGs that are common to the *ancient eight* species, which comprise the *ancient six* plus *Aeropyrum pernix* (Ape) and *Pyrobaculum aerophilum* (Pae), the two primitive free living crenarchaeons with the lowest alloacceptor tRNA distances (5). These 424 genes give a minimum representation of the LUCA proteome. Combining this minimum LUCA proteome with, based on the Mka genome, 39 structural RNA genes yields a minimum LUCA genome of 463 genes, which is far smaller than the Mka genome of 1731 genes (10). So either genome size has evolved substantially between LUCA and Mka, or the 463 genes formed only a portion of the LUCA genome, or likely both. An estimate of over 1000 LUCA genes based on ancestral state inference of gene content (11) is in agreement with the 463 minimum LUCA genes being only a portion of the LUCA genome.

Whether LUCA was a rudimentary progenote or a full-fledged genote has been one of the key unanswered evolutionary questions (12). The size of the minimal proteome compatible with life is given by models I-VI, which in increasing order of the estimated proteome size consist of 150-340 protein encoding genes:

I. Genes in a "limping" life form (13): 150 genes

II. The minimalist genome of *Carsonella ruddii* (14): 182 genes

III. Genes in the hypothetical minimal cell (15): 200 genes

IV. Minimal proteome deduced from a comparison of *Mycoplasma genitalium* and *Haemophilus influenzae* genomes (16): 256 genes.

V. *Bacillus subtilis* essential genes identified by deletions (17): 271 genes

VI. Core protein genes of bacteria (18): 340 genes

Since the minimum LUCA proteome of 424 COGs exceeds significantly the minimal proteomes I-VI of150-340 genes assessed to be compatible with life, a range of "non-minimal" genes formed part of the LUCA genome. For instance, the COGs belonging to groups (A) RNA processing and modification, (B) chromatin structure and dynamics, (D) cell cycle control, cell division, chromosome partitioning, and (T) signal transduction mechanisms (9) are part of the LUCA proteome in Appendix 1 based on either the *ancient six* or the *ancient eight*, but are not included in any of the minimal proteomes I-VI. *Consequently LUCA was not a progenote or a minimal cell*. Evidently, being the last prebiotic organism and the first modern organism, LUCA had evolved beyond the minimal cell stage of life. Support for a genetically complex LUCA has also been obtained from the analysis of protein families (19, 20).

5. MISSING GENES

Biological evolution is characterized by the continual introduction of novel genes through such processes as gene duplication and exon shuffling both prior and subsequent to the emergence of LUCA. All those genes that were introduced into LUCA's descendents would be missing from LUCA. The human genome contains some 20,000 genes, and some organisms have even larger genomes. Accordingly the great majority of genes found in the biological world were missing from the LUCA genome. These missing genes testify to the tremendous growth of gene space among LUCA's descendants as they adapt to all kinds of ecological niches on Earth. Some examples of missing genes are as follows:

5.1. Gln-, Asn- and Cys-tRNA synthetase genes

The genetic code evolved from a primitive code that encoded about ten primordial Phase 1 amino acids readily obtainable from prebiotic synthesis or meteorites to the present day twenty standard amino acid code comprising the Phase 1 amino acids as well as Phase 2, biosynthetically-derived amino acids. Some of the Phase 2 amino acids were produced at first by pretran synthesis, which converts a Precursor aa-tRNA compound to a Product aa-tRNA compound. Since the conversion is accomplished while the precursor stays attached to the tRNA, the nascent product amino acid acquires the anticodon on the tRNA as soon as it is formed, thereby assuring its entry into the genetic code. Even to-day, the pretran synthesis pathway still operates in a variety of organisms to produce the Gln-tRNA, Asn-tRNA and Cys-tRNA compounds to serve the incorporation of Gln, Asn and Cys residues into proteins. In other organisms, the pretran synthesis pathway has been superseded by GlnRS, AsnRS or CysRS, which joins Gln, Asn or Cys directly to its cognate tRNA. As to be expected from the late-comer status of GlnRS, AsnRS and CysRS (8,21, 22), the genes for these three synthetases are missing from the LUCA genome (Appendix 1).

5.2. Cytochrome genes

The biological world is opportunistic and utilizes a wide spectrum of reductant-oxidant combinations to generate ATP through coupling to electron transport (23). Cytochromes are widely employed electron carriers participating in this task. However, when LUCA first arose in the hydrothermal vents, its strongly electro-negative H_2-CO_2 reductant-oxidant combination functioned outside the usual range of cytochrome mid-point potentials. Therefore it is not surprising that cytochrome genes were absent from the LUCA genome (8, and Appendix 1).

6. LATE ARRIVAL OF DNA GENOMES

The minimum LUCA proteome derived from the *ancient eight* encodes 24 COGs for DNA replication, recombination and repair, which establishes that LUCA possessed a DNA genome. The genes from the three biological domains relating to DNA informational molecules are known to display erratic species diversity (24). For example, the topoisomerase II (also named IIA) gene is present in some euryarchaeons, but absent from the crenarchaeons and also other euryarchaeons, which had seemed to be a rather puzzling distribution. It turns out that the *ancient eight* archaeons all lacking topoisomerase II are in fact located close to LUCA on the tRNA tree, either on the *Crenarchaeota* side or on the *Euryarchaeota* side. Therefore, because LUCA lacked topoisomerase II (Appendix 1), the enzyme is absent from those archaeons close to LUCA, but present in some of the archaeons more distant from LUCA. This way, the distribution pattern for the enzyme is explained by the distance from LUCA. Previously, it was proposed that RNA genomes preceded DNA genomes (25). The use of a Cys-radical in ribonucleotide reductase that was attainable only with enzymes but not with ribozymes further suggests that DNA genomes were preceded by proteins (26). The lack of topoisomerase II in LUCA, and the erratic species diversity of DNA informational genes are both in accord with such evidence for the late arrival of DNA, resulting in the DNA informational system still undergoing fundamental adaptations at and after the LUCA stage of life.

The displacement of RNA genomes by DNA genomes could be caused by: (i) the greater stability of the

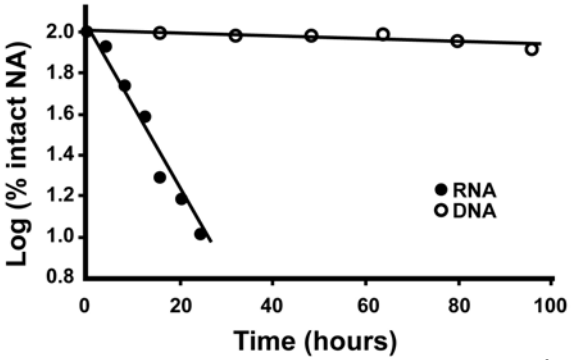

Figure 1. Thermal degradation of RNA and DNA at 100°C. The RNA sequence employed was *gauucaaucugaucucgaugaag*, and the corresponding DNA sequence employed was *gattcaatctgatctcgatgaag*. The buffer was 0.05M potassium phosphate and 0.05M sodium phosphate, pH7. Degradation of the 23-mers was monitored by gel electrophoresis with band quantitation by GelDoc. The estimated $t_{1/2}$ was 7.5 hours for RNA and 501 hours for DNA.

phosphodiester backbone of DNA compared to RNA; (ii) absence of proofreading by RNA polymerases leading to highert mutation rates in RNA genomes; (iii) information in RNA degrades because deamination of cytosine to form uracil can be detected and repaired in DNA but not in RNA, thereby leading to higher error rates in RNA genomes compared to DNA genomes; (iv) UV irradiation produces more photochemical changes in RNA than in DNA (25). Consequently DNA genomes are more advantageous than RNA genomes. Yet the parallel existence of RNA and DNA viruses in nature suggests that it is an open question whether or not the identified selective advantages of DNA over RNA might be strong enough to drive the development of the highly intricate ensemble of DNA informational molecules. This uncertainty has led to the suggestion that the rise of DNA organisms was brought about by the need of RNA organisms to defend against attack by RNA viruses (24,27).

However, the selective advantage of DNA genomes over RNA genomes is far from modest at elevated temperatures. As Figure 1 shows, the chemical half-life of RNA at hyperthermophilic temperature is so shortened that RNA genomes would lose viability. Consequently the transition from RNA genomes to DNA genomes would constitute not a mere preference but an absolute necessity at the hyperthermophilic temperatures. In keeping with this, there are few if any RNA viruses that infect the archaeal hyperthermophiles (28). Adaptation to hyperthermophilic conditions therefore could provide a pivotal evolutionary incentive for the development of DNA genomes, allowing the hyperthermophiles to thrive at the hydrothermal vents, and paving the way to the ascendance of a hyperthermophilic LUCA.

7. ORIGIN OF LIFE SCENARIOS

What kind of temperature scenario for the origin of life is in accord with a hyperthermophilic LUCA? Enquiries into life's origin, starting from the first self-replicating systems and leading up to LUCA, have revealed varying effects of different thermal environments on the process. The identification of a hyperthermophilic LUCA raises the question of how the emergence pathway for early life forms was shaped by these thermal effects to reach the hydrothermal vents. In this regard, the different thermal environments may be referred to in terms of the organisms they host, be they psychrophiles, mesophiles, thermophiles or hyperthermophiles.

7.1. Mesophilic Origin (MEO)

Just as the majority of known organisms are mesophiles, many origin of life scenarios consider a milieu consisting of bodies of surface water on Earth, exemplified by Darwin's "warm little pond"(29), that are inhabited to-day by mesophiles.

Advantage: In support of MEO, there is plentiful evidence for the prebiotic synthesis of organic compounds in the atmosphere, and derivation of organic compounds from meteorites, hydrothermal vents and mineral/clay catalysis, together providing a heterotrophic basis for the start of mesophilic life feeding on the accumulated organic compounds in the environment (30).

Disadvantage: The proliferation of organisms knows no bounds, relentlessly driving toward exhaustion of resources – this was as true in prebiotic times as to-day. Since glycolysis leading from one glucose to two lactate molecules reduces five C-C bonds to four C-C bonds, and that leading from one glucose to two ethanol molecules reduces five C-C bonds to two C-C bonds, the heterotrophic world is dependent on continuous carbon fixation to form covalent C-C bonds. Nowadays photosynthesis supplies ample carbon fixation. However, between life's emergence before 3 Gya and the advent of photosynthesis at about 2.3 Gya, the heterotrophic world would sooner or later encounter or approach a *carbon-fixation crisis* as its exponential expansion outpaced the linear production of environmental organics. Facing this crisis, organisms could turn to chemolithotrophy for energy, but they also had to secure carbon fixation through chemoautotrophy. Since few organisms have been able to achieve efficient non-photosynthetic chemoautotrophy under mesophilic conditions, the survival and expansion of the MEO living world could be severely limited by the carbon-fixation crisis prior to the invention of photosynthesis.

7.2. Psychrophilic Origin (PSO)

Advantage: Because solar irradiation was less intense on primitive Earth under a faint Sun, water-ice mixtures could be found in various regions. While similar to MEO, PSO offers the advantage of a more facile condensation of RNA monomers in the presence of ice nucleation, e.g. when the four imidazole-activated monoribonucleotides are incubated with Mg(II)/Pb(II) mixtures at slightly below the freezing point, up to 90% quasi-equimolar incorporation of all the monomers into 5 to 17-mers is observed with traces of longer products (31). Since the prebiotic synthesis of RNA or RNA-like

oligomers and polymers represents one of the key barriers in life's emergence, this PSO advantage could be an important one.

Disadvantage: PSO has the same disadvantage as MEO. In addition, reaction rates will be slower at the lower temperatures.

7.3. Hyperthermophilic Origin (HYO)

Ever since the discovery of living communities at the submarine hydrothermal vents, these vents have been proposed as sites for the origin of life (32). Under the HYO scenario, the geothermal energy released at these vents could provide a hospitable environment for the origin of life. Moreover, these vents could be more widespread on early Earth than to-day.

Advantage: The hydrothermal vents are endowed with abundant thermal energy, as well as carbon dioxide and hydrogen that could be captured for lithoautotrophy, averting any prospect of a carbon-fixation crisis prior to the invention of photosynthesis. Carbon fixation is exergonic at the elevated vent temperatures (33), and prebiotic synthesis under vent-like conditions has been demonstrated for a range of organic compounds (34-37).

Furthermore, large asteroid impacts on early Earth posed a threat to the primordial biota. It is estimated that the impact of a 440-km diameter (1.3×10^{20} kg) projectile about the size of Vesta and Pallas might bring the oceans to a boil or near boil (38). Such calamities could obliterate mesophilic life but leave the hyperthermophiles viable.

Disadvantage: A forbidding drawback of HYO is the instability at elevated temperatures of various organic compounds essential to life's emergence, which suggests that the hyperthermophiles could not be the first life forms but had to be derived through secondary adaptation from inhabitants of cooler zones (39-41). In addition, some metabolic intermediates are labile even at cooler temperatures, managing to fulfill their metabolic roles in extant organisms only through metabolite channeling where a labile intermediate is protected against degradation by protein shielding. Since the shielding depends on evolved enzyme, or in earlier times ribozyme, structures, metabolic channeling would be difficult to accomplish at the very beginning of biotic evolution when the biocatalysts were unsophisticated in their structures. Furthermore, prebiotic RNA or RNA-like replication depends on efficient base pairing and base stacking, processes that are easily disrupted by high temperatures.

As well, to evolve a novel enzyme function from a pre-existing protein by gene duplication or exon shuffling under hyperthermophilic conditions, it is necessary to achieve mutationally at once the new catalyst function and the stabilization of the new catalyst conformation, a daunting task at elevated temperatures. This difficulty in generating novel enzyme functions might well contribute to the ultra-conservative, slow-evolving nature of the hyperthermophiles, resulting in their phylogenetic placements at the roots of the Archaea and Bacteria domains (42). It further strengthens the suggestion that hyperthermophiles originally evolved from inhabitants of cooler zones in contradiction to HYO (41).

7.4. Thermophilic Origin (THO)

Advantage: To ameliorate the chemical instabilities encountered at hyperthermophilic temperatures, and still capture the biosynthetic advantages of the hydrothermal vents, it has been proposed that life could have originated not at the vents themselves, but close to the vents at <90^{o}C (35,43). Such a THO scenario could offer a more favorable balance between the chemosynthetic benefits and the perils of chemical instabilities at the vents.

Disadvantage: The melting temperature for the binding of a DNA 20-mer consisting of half purines and half pyrimidines to template DNA is only about 52^{o}C. So template-directed condensation of mononucleotides into RNA, RNA-like or DNA oligomers much shorter than a 20-mer in length would be difficult to initiate prebiotically at either the thermophilic or hyperthermophilic temperatures.

7.5. Hot Cross Origin (HCO)

In view of the debilitating shortcomings of the MEO, PSO, HYO and THO scenarios, the identification of a hyperthermophilic *Methanopyrus*-proximal LUCA has led to the proposal of a Hot Cross Origin of life (44), which suggests that life began at the psychrophilic to mesophilic temperature zones where template-directed synthesis of RNA or RNA-like oligomers to polymers was facilitated. As the heterotrophic life forms multiplied endlessly and a carbon-fixation crisis threatened, exploitation of new sources of organic compounds such as those arising from carbon fixation at the hydrothermal vents became important. The emergent biota from lower temperatures adapted at first to the thermophilic zone in the vicinity of these vents, and later on as competition intensified to the hyperthermophilic conditions at the vents, by thermal proofing their enzymes, inventing the DNA genome, and developing methanogenesis on the foundation of heterotrophic metabolic pathways (45). Thus the life forms crossed progressively from the lower temperature zones into the hyperthermophilic zone. At the vents, the methanogens flourished on the plentiful CO_2 and H_2, producing massive amounts of methane to ward off an ice age even under a faint Sun (46), finalized the 20-amino acid universal genetic code, and perfected a DNA genome to give rise to LUCA. Later on, the swarms of LUCA descendents would re-cross the temperature zones in the opposite direction, adaptively radiating back to the thermophilic, mesophilic and psychrophilic environments. Armed with the universal genetic code and a DNA genome, they overran and eliminated all competitor organisms remaining in those environments which, never having undergone adaptation to hyperthermophlic temperatures, were still equipped with an error-prone RNA genome that was no match against the newly developed DNA genome.

Advantage: HCO explains the rise of a hyperthermophilic, methanogenic LUCA and along with it

the universal genetic code. The scenario also identifies adaptation to hydrothermal vent temperatures as a powerful enough selection factor to drive the transition from RNA genomes to DNA genomes, which required the elaboration of a multi-enzyme DNA informational machinery. Furthermore, while carbon fixation crises and near global sterilizations would accelerate migration to the vents, HCO was not inherently dependent on either of these occurrences. The rich but scaldingly hot vents supported the vast proliferation of life forms adapted to the vents, and their perfection of a heat-resistant DNA genome, which together with the universal genetic code ensured the ascendance of a hyperthermophilic methanogenic LUCA from their ranks. LUCA's descendants, nurtured to unprecedented numbers and armed with the new biochemical weaponry of perfected DNA genome and genetic code, became matchless when they fanned out and invaded the lower temperature realms where the life forms were decimated by carbon fixation crisis and saddled with an RNA genome and likely an inferior genetic code.

Disadvantage: Species adaptations to enhanced temperature resistance are commonplace, and HCO is free of any obvious disadvantage.

8. DISCUSSION

The origin of life began with the first replicating informational molecules and culminated in the emergence of LUCA. The finding of a LUCA closely related to the hyperthermophilic archaeal methanogen *Methanopyrus* opens the door to an examination of the biology and evolution of LUCA. The elucidation of the minimum LUCA proteome based on the genes of the *ancient eight* has revealed a LUCA that was no longer a minimal organism. It was instead a simple but full-fledged modern organism. The modernity of its molecular biology and biochemistry, including its DNA genome and universal genetic code, accounts for the unparalelled dominance of its lineage over all its contemporaries (21,22,47).

The presence of the multiple COGs for DNA replication, recombination and repair in its genome indicates that LUCA utilized a DNA genome. However, LUCA's lack of topoisomerase II (Appendix 1), the erratic species diversity of DNA informational genes (24), and the dependence on a Cys-radical by ribonucleotide reductase to synthesize deoxyribonucleotides (26) all suggest that the DNA genome was a late evolutionary development that arrived not long before the rise of LUCA. Earlier it was suggested that the erratic species diversity of DNA informational genes might be traced to the separate transfers of these genes into the individual biological domains by viruses (24,27). However, it is difficult to determine the direction of horizontal gene transfers between the viral and cellular lineages, and the viruses could have acquired their genes by ancient horizontal gene transfers from extinct cellular lineages in the first place (48). Furthermore, erratic species diversity is also observed in the pathways and enzymes of archaeal carbohydrate metabolism, where the existence of a wide variety of alternate enzymes often from different enzyme families points to a greater metabolic diversity in Archaea compared to Bacteria and Eukarya (49). Since viruses seldom encode carbohydrate metabolic enzymes, they were unlikely to be the cause of carbohydrate metabolic diversity in Archaea, which was more easily explained by post-LUCA developments of non-methanogenic carbohydrate metabolism by many of LUCA's offsprings as they abandoned methanogenesis and spread out from the hydrothermal vents into cooler zones. Likewise, the erratic species diversity of DNA informational genes observed in the biological domains might have been caused by the late arrival of DNA genomes not long before the rise of LUCA and the separation of the Archaea, Bacteria and Eukarya domains not too long after the LUCA stage, as described in the Hot Cross Origin.

The MEO, PSO, HYO and THO scenarios for the origin of life are all burdened with crippling difficulties. Moreover, in order for MEO, PSO and THO to be compatible with a hyperthermophilic LUCA, evolutionary factors have to be envisioned that would shift the center of gravity of life from the cooler zones to the hyperthermophilic zone, be it a carbon fixation crisis, a near global sterilization event, a combination of both, or some other factors. Regardless of the actual cause or causes, such a shift in effect converts each of these scenarios into the Hot Cross Origin scenario. As well, the origin of life is astounding not only in its occurrence but also in its rapidity. Life appeared on primordial Earth within several hundred million years, whereas it has taken three billion years to evolve vertebrate intelligence. In contrast to the inherent hurdles and dead-ends encountered by the other scenarios, the Hot Cross Origin combines all the requisite elements for a rapid emergence of life on Earth. By enlisting the initiation of RNA or RNA-like replication at psychrophilic to mesophilic temperatures, the practice of heterotrophy early on making good use of the environmental organic compounds derived from prebiotic syntheses and meteorites, capabilities to survive carbon-fixation crises and near global sterilizations by asteroid impacts, as well as a metabolic rationale for inventing the DNA genome, it provides a facilitated pathway to bring the early life forms to a LUCA at the hydrothermal vents.

One long standing riddle regarding the origin of life is the pathway by which a single LUCA lineage came to entrench its genetic code as the universal code, eliminating in the process all other contemporary lineages bearing alternate codes (47). The mystery only deepens in the face of the overwhelming evidence for a hyperthermophilic methanogenic LUCA (8), for lineages inhabiting different temperature zones do not compete directly and there is no apparent mechanism by which a hyperthermophile could kill off a mesophile or psychrophile. The Hot Cross Origin finally provides an answer: first, early life forms adapted to the hydrothermal vents on account of the benefit of lithoautotrophy offered by methanogenesis; secondly, in order to inhabit the vents and tap into the riches of methanogenesis, they had no choice but to construct an elaborate DNA informational machinery and switch from an RNA genome to a DNA genome; at the end it was likely the DNA genome

fashioned at the crucible of the vents, which because of its lower mutation and error rates could function more efficiently and permit larger genome sizes than the RNA genome, that together with the perfected genetic code itself that proved to be the unsparing biochemical weapons enabling LUCA's descendants to eliminate all competitor lineages in the cooler zones that lived by alternate genetic codes and an RNA genome when LUCA's descendants radiated back to those zones. The resultant absolute conquest of the entire living world by LUCA's descendants established the universality of the present day 4-letter DNA genome and 20-letter genetic code languages of cellular life, just as the Roman Empire established the Latin languages throughout Europe by means of all-conquering military might.

Directed by natural selection, migratory birds cross continents yearly to search for food, and LUCA's forebears and offsprings crossed temperature zones repeatedly for survival, proliferation and unknowing conquest.

9. ACKNOWLEDGEMENT

We thank the Research Grants Council of Hong Kong (Grant No. HKUST6437/06M) for support.

10. REFERENCES

1. Elizabeth Pennisi: Is it time to uproot the tree of life? *Science* 284, 1305-1307 (1999)

2. W. Ford Doolittle: Phylogenetic classification and the universal tree. *Science* 284,2124-2129 (1999).

3. Yuri I. Wolf, Igor B. Rogozin, Nick V. Grishin, Eugene V. Koonin: Genome trees and the tree of life. *Trends Gene* 18, 472-479 (2002).

4. J. Tze-Fei Wong, Hong Xue: Self-perfecting evolution of heteropolymer building blocks and sequences as the basis of life. In: Fundamentals of Life. Eds: Palyi G, Zucchi C, Caglioti L. Elsevier, Paris p. 473-494 (2002).

5. Hong Xue, Ka-Lok Tong, Christian Marck, Henri Grosjean, J. Tze-Fei Wong: Transfer RNA paralogs: evidence for genetic code-amino acid biosynthesis coevolution and an archaeal root of life. *Gene* 310, 59-66 (2003)

6. Ka-Lok Tong, J. Tze-Fei Wong: Anticodon and wobble evolution. *Gene* 333, 169-177 (2004)

7. Hong Xue, Siu-Kin Ng, Ka-Lok Tong, J. Tze-Fei Wong: Congruence of evidence for a *Methanopyrus*-proximal root of life based on transfer RNA and aminoacyl-tRNA synthetase genes. *Gene* 360, 120-130 (2005)

8. J. Tze-Fei Wong, Jianhuan Chen, Wai-Kin Mat, Siu-Kin Ng, Hong Xue: Polyphasic evidence delineating the root of life and roots of biological domains. *Gene* 203, 39-52 (2007)

9. Roman L. Tatusov, Eugene V. Koonin, David J. Lipman: a genomic perspective on protein families. *Science* 278, 631-637 (1997).

10. Alexei Slesarev *et al*: The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci USA* 99, 4644-4649 (2002)

11. Christos A. Ouzounis, Victor Kunin, Nikos Darzentas, Leon Goldovsky: A minimum estimate for the gene content of the last universal common ancestor—exobiology from a terrestrial perspective. *Res Micobiol* 157, 57-68 (2006).

12. Carl R. Woese: Evolutionary questions: the "Progenote". *Science* 247, 789 (1990)

13. Markus Pech and Knud H. Nierhaus. The minimal cell. In: Prebiotic Evolution and Astrobiology. Eds: J. Tze-Fei Wong, Antonio Lazcano. Landes Bioscience USA (2008) in press.

14. Atsushi Nakabachi, Atsushi Yamashita,Hidehiro Toh, Hajime Ishikawa, Helen E. Dunbar, Nancy A. Moran, Masahira Hattori: The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314, 267 (2006)

15. Sara Islas, Arturo Becerra, P. Luigiluisi, Antonio Lazcano: Comparative genomics and the gene complement of a minimal cell. *Orig Life Evol Biosph* 34, 243-56 (2004)

16. John I. Glass, Macyra Assad-Garcia, Nina Alperovich, Shibu Yooseph, Matthew R. Lewis, Mahir Maruf, Clyde A. Hutchison III, Hamilton O. Smith, J. Craig Venter: Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* 103, 425-430 (2006)

17. K. Kobayashi *et al*: Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci U S A* 100, 4678-4683 (2003)

18. Antoine Danchin, Gang Fang, Stanislas Noria: The extant core bacterial proteome is an archive of the origin of life. *Proteomics* 7, 875-889 (2007)

19. Juan A. Ranea, Antonio Sillero, Janet M. Thornton, Christine A. Orengo: Protein superfamily evolution and the Last Universal Common Ancestor (LUCA). *J Mol Evol* 63, 513-525 (2006)

20. Minglei Wang, Liudmila S. Yafremava, Derek Caetano-Anolles, Jay E. Mittenthal, Gustavo Caetano-Anolles: Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. *Genome Res* 17, 1572-1585 (2007)

21. J. Tze-Fei Wong: Coevolution theory of the genetic code at age thirty. *BioEssays* 27, 416-425 (2005)

22. J. Tze-Fei Wong: Question 6: Coevolution theory of the genetic code: a proven theory. *Orig Life Evol Biosph* 37, 403-408 (2007)

23. Kenneth H. Nealson, Pamela G. Conrad: Life: past, present and future. *Phil Trans R Soc Lond B* 354, 1923-1939 (1999)

24. Patrick Forterre: Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. *Proc Natl Acad Sci U S A* 103, 3669-3674 (2006)

25. A. Lazcano, R. Guerrero, L. Margulis, J. Oro: The evolutionary transition from RNA to DNA in early cells. *J Mol Evol* 27, 283-290 (1988).

26. Stephen J. Freeland, Robin D. Knight, Laura F. Landweber: Do proteins predate DNA? *Science* 286, 690-692 (1999).

27. Patrick Forterre: The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* 87,793-803 (2005)

28. David Prangishvili, Patrick Forterre, Roger A. Garrett: Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* 4,837-848 (2006).

29. F. Darwin: Letter to Joseph Dalton Hooker, Feb. 1, 1871 in The Life and Letters of Charles Darwin, vol. II. (1959) Basic Books, New York, pp. 202-203.

30. Antonio Lazcano, Stanley L. Miller: On the origin of metabolic pathways. *J Mol Evol* 49, 424-31 (1999)

31. Pierre-Alain Monnard, Anastassia Kanavarioti, David W. Deamer: Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J Am Chem Soc* 125, 13734-13740 (2003)

32. John. B. Corliss, John A. Baross, Sarah E. Hoffman: An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. Oceanol. Acta 4, 59-69. Proc Int Geol Congress Geol Oceans Symp. Paris, France (1981).

33. Everett L. Shock: Chemical environment in submarine hydrothermal systems. *Orig Life Evol Biosph* (suppl.) 22, 67-107 (1992)

34. N. G. Holm, E.M. Andersson: Abiotic synthesis of organic compounds under the conditions of submarine hydrothermal systems: a perspective. *Planet Space Sci* 43, 153-159 (1995).

35. George D. Cody, Nabil Z. Boctor, Timothy R. Filley, Robert M. Hazen, James H. Scott, Anurag Sharma, Hatten S. Yoder Jr: Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate. *Science* 289, 1337-1340 (2000)

36. Claudia Huber, Günter Wächtershauser: Alpha-Hydroxy and alpha-amino acids under possible Hadean, volcanic origin-of-life conditions. *Science* 314,630-632 (2006)

37. Tomas M. McCollom, Jeffrey S. Seewald: Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Chem Rev* 107, 382-401 (2007)

38. Norman H. Sleep, Kevin J. Zahnle, James F. Kasting, Harold J. Morowitz: Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature* 342, 139-142 (1989)

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

40. Norman R. Pace: Origin of life--facing up to the physical setting. *Cell* 65, 531-533 (1991)

41. Sara Islas, Ana M. Velasco, Arturo Becerra, Luis Delaye, Antonio Lazcano: Hyperthermophily and the origin and earliest evolution of life. *Int Microbiol* 6, 87-94 (2003)

42. Karl O Stetter: Hyperthermophilic prokaryotes. *FEMS Microbiol Rev* 18, 149-158 (1996)

43. Eugene V. Koonin, William Martin: On the origin of genomes and cells within inorganic compartments. *Trends Genet* 21, 647-654 (2005)

44. J. Tze-Fei Wong: The root of life. In: Prebiotic Evolution and Astrobiology. Eds: J. Tze-Fei Wong, Antonio Lazcano. Landes Bioscience USA (2008).

45. James G. Ferry, Christopher H. House: The stepwise evolution of early life driven by energy conservation. *Mol Biol Evol* 23, 1286-1292 (2006)

46. James F. Kasting, Janet L. Siefert: Life and the evolution of Earth's atmosphere. *Science* 296, 1066-1068 (2002)

47. J. Tze-Fei Wong: The evolution of a universal genetic code. *Proc Natl Acad Sci U S A* 73, 2336-40 (1976)

48. Patrick Forterre, Simonetta Gribaldo, Daniele Gadelle, Marie-Claude Serre: Origin and evolution of DNA topoisomerases. *Biochimie* 89, 427-446 (2007)

49. Bettina Siebers and Peter Schönheit: Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr Opin Mirobiol* 8, 695-705 (2005)

Key Words : Last Universal Common Ancestor, LUCA, minimal cell, Hot Cross Origin, Review

Send correspondence to: J. Tze-Fei Wong, Department of Biochemistry and Applied Genomics Center, Hong Kong University of Science and Technology, Hong Kong, China. Tel: 852-2358-7288, Fax: 852-2358-1552, E-mail: bcjtw@ust.hk

http://www.bioscience.org/current/vol13.htm