REVIEW ARTICLE

Sara Islas · Ana M. Velasco · Arturo Becerra
Luis Delaye · Antonio Lazcano

Hyperthermophily and the origin and earliest evolution of life

Abstract The possibility of a high-temperature origin of life has gained support based on indirect evidence of a hot, early Earth and on the basal position of hyperthermophilic organisms in rRNA-based phylogenies. However, although the availability of more than 80 completely sequenced cellular genomes has led to the identification of hyperthermophilic-specific traits, such as a trend towards smaller genomes, reduced protein-encoding gene sizes, and glutamic-acid-rich simple sequences, none of these characteristics are in themselves an indication of primitiveness. There is no geological evidence for the physical setting in which life arose, but current models suggest that the Earth's surface cooled down rapidly. Moreover, at 100 °C the half-lives of several organic compounds, including ribose, nucleobases, and amino acids, which are generally thought to have been essential for the emergence of the first living systems, are too short to allow for their accumulation in the prebiotic environment. Accordingly, if hyperthermophily is not truly primordial, then heat-loving lifestyles may be relics of a secondary adaptation that evolved after the origin of life, and before or soon after separation of the major lineages.

Keywords Hyperthermophily · Comparative genomics · Organic-compound stability · Last common ancestor · Origin of life
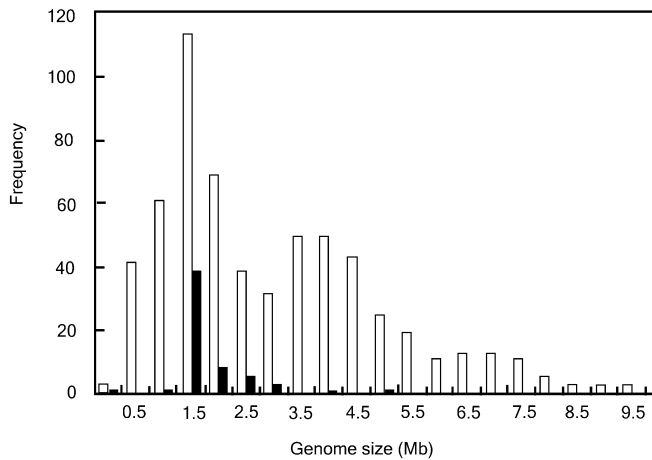
Introduction

A thermophilic origin of life is not a new idea. "Heat has been justly regarded the mother of all generations," wrote Jean-Baptiste Lamarck in his 1804 *Philosophie Zoologique*, adding that "it cannot be doubted that suitable portions of inorganic matter, occurring amidst favorable surroundings, may by the influence of Nature's agents, of which heat and moisture are the chief, receive an arrangement of their parts that foreshadows cellular organization, and thereafter pass to the simplest organic state and manifest the earliest movements of life" [22].

Lamarck's ideas are echoed in a number of contemporary proposals on a hot origin of life. It is not surprising that the correlation between hyperthermophily and antiquity has led to suggestions of a high-temperature emergence of life. This interpretation has been reinforced by a number of facts, including large-scale analysis suggesting that, soon after its formation, the surface of primitive Earth was extremely hot. The planet is generally thought to have remained molten for some time after its formation 4.6×10^9 years ago, although evidence of a 4.4×10^9-year-old hydrosphere implies that its surface cooled down rapidly [52]. However, the Earth underwent late accretion impacts that may have boiled-off the oceans as late as 3.8×10^9 years ago [41]. Moreover, both paleontological and molecular fossil records appear to support the possibility of a hyperthermophilic origin of life: (a) the 3.49- to 3.43×10^9-year-old Australian Warrawoona stromatolitic chert horizons [37] are endowed with the diagnostic features of a microbial community associated with a seafloor hydrothermal system [49]; and (b) rooted universal single-gene phylogenies have shown that hyperthermophiles are not randomly distributed in universal trees, but occupy the deepest, shorter branches towards the lowest portion of molecular rRNA-based cladograms [1, 34, 44].

However, attempts to infer the conditions of life based on the traits of heat-loving prokaryotes have led to opposing suggestions: while some advocate a hyperthermophilic heterotrophic emergence of life [7, 19], others hypothesize that mineral surfaces in hot volcanic settings fueled the appearance of primordial chemo-autolithotrophic biological systems lacking genetic material [50]. Regardless of these differences, all

S. Islas · A. M. Velasco · A. Becerra · L. Delaye · A. Lazcano (✉)
Facultad de Ciencias, UNAM, Ciudad Universitaria,
Apdo. Postal 70-407, 04510 México D.F., México
E-mail: alar@correo.unam.mx
Tel.: +52-5-6224823
Fax: +52-5-6224828

Fig. 1 Chromosomal DNA content distribution for mesophiles (white bars) and hyperthermo- and thermophilic (growth temperature 80 °C or more) prokaryotes (black bars). Prokaryotic genome sizes shown here were reported in the literature as determined by pulsed-field gel electrophoresis ($n = 641$)

Fig. 2 Protein-encoding gene size distribution as a function of genome size. *Black dots* correspond to extremophilic proteomes of species that live at 80 °C or more. The area within which extremophilic ORFs are located is *circled* for clarity. *Symbols*: Archaea hyperthermophilic crenarchaea: *Aeropyrum pernix* (App), *Sulfolobus solfataricus* (Sus), *Sulfolobus tokodaii* (Sut), *Pyrobaculum aerophilum* (Pya); hyperthermophilic euryarchaea: *Pyrococcus abyssi* (Pyb), *Pyrococcus furiosus* (Pyf), *Pyrococcus horikoshii* (Pyh), *Methanococcus jannaschii* (Mej), *Thermoplasma acidophilum* (Tha), *Methanobacterium thermoautotrophicum* (Met), *Archaeoglobus fulgidus* (Arf), *Thermoplasma volcanium* (Thv); non-hyperthermophilic euryarchaea: *Halobacterium* sp. (Has), *Methanopyrus kandleri* (Mek), *Methanosarcina mazei* (Mem), *Methanosarcina acetivorans* (Mea); Bacteria hyperthermophilic bacteria: *Aquifex aeolicus* (Aqa), *Thermotoga maritima* (Thm); α-Proteobacteria: *Rickettsia conorii* (Ric), *Rickettsia prowazekii* (Rip), *Sinorhizobium meliloti* (Sim1, Sim2, Sim3), *Brucella melitensis* (Brm, Brm2), *Agrobacterium tumefaciens* C58 Uwash (AgtU), *Agrobacterium tumefaciens* C58 Cereon (AgtC, AgtC2), *Mesorhizobium loti* (Mel), *Caulobacter crescentus* (Cac); Beta proteobacteria: *Neisseria meningitidis* MC58 (Nem8a, Nem8b), *Ralstonia solanacearum* (Ras, Ras2); γ-Proteobacteria: *Xylella fastidiosa* (Xyf), *Vibrio cholerae* (Vic, Vic2), *Haemophilus influenzae* (Hai), *Salmonella typhi* (Sat), *Salmonella typhimurium* LT2 (Sat2), *Escherichia coli* K12 (Eco2), *Escherichia coli* O157H7 (Eco7), *Escherichia coli* O157H7 EDL933 (Eco3), *Yersinia pestis* KIM (YepM), *Yersinia pestis* CO92 (Yep2), *Pseudomonas aeruginosa* (Psa), *Xanthomonas citri* (Xac), *Xanthomonas campestris* (Xaa), *Pasteurella multocida* (Pam), *Buchnera aphidicola* Sg (Bua), *Buchnera* sp. (Bus);δ/ε-Proteobacteria: *Campylobacter jejuni* (Caj), *Helicobacter pylori* 26695 (Hep5), *Helicobacter pylori* J99 (Hep9); green sulfur bacteria: *Chlorobium tepidum* TLS (Cht); gram-positive, low-GC: *Streptococcus pneumoniae* TIGR4 (Stp4), *Streptococcus pneumoniae* R6 (Stp6), *Streptococcus pyogenes* MGAS315 (Sty5), *Listeria innocua* (Lii), *Listeria monocytogenes* (Lim), *Thermoanaerobacter tengcongensis* (Tht), *Staphylococcus aureus* MW2 (Sta2), *Staphylococcus aureus* Mu50 (Sta0), *Staphylococcus aureus* N315 (Sta5), *Lactococcus lactis* (Lal), *Streptococcus pyogenes* (Sto), *Clostridium perfringens* (Clp), *Clostridium acetobutylicum* (Cla), *Bacillus subtilis* (Bsu), *Bacillus halodurans* (Bah), *Mycoplasma pneumoniae* (Myn), *Mycoplasma genitalium* (Myg), *Mycoplasma pulmonis* (Myp), *Ureaplasma urealyticum* (Uru); gram-positive, high-GC: *Streptomyces coelicolor* (Stc), *Mycobacterium tuberculosis* CDC1551 (Myt1), *Mycobacterium tuberculosis* H37Rv (Mytv), *Mycobacterium leprae* (Myl); radioresistant bacteria: *Deinococcus radiodurans* (Der1, Der2); Fusobacterium: *Fusobacterium nucleatum* (Fun); cyanobacteria: *Synechocystis* CC6803 (Syn), *Thermosynechococcus elongatus* (The), *Nostoc* sp. (Nos); actinobacteria: *Corynebacterium glutamicum* (Cog); chlamydia: *Chlamydophila pneumoniae* AR39 (Chp9), *Chlamydophila pneumoniae* J138 (Chp8), *Chlamydophila pneumoniae* CWL029 (Chp2), *Chlamydia trachomatis* (Chr), *Chlamydia muridarum* (Chm); spirochete: *Borrelia burgdorferi* (Bob), *Treponema pallidum* (Trp)

hot-origin-of-life scenarios face the same problem, i.e., the chemical decomposition of presumed essential biochemical compounds, such as amino acids, ribose, nucleobases, RNA, and other thermolabile molecules, whose half-lives for decomposition at temperatures between 250 °C and 350 °C are at the most a few minutes [28, 51].

Is it possible, then, that the evidence supporting a hot origin of life is being misinterpreted, i.e., that the extrapolation of molecular phylogenies into prebiotic times is misleading? The purpose of this paper is to review the evidence against an extremophilic origin of life and a heat-loving RNA world, thus supporting the possibility that (hyper) thermophilic microbial lifestyles are in fact the outcome of secondary adaptations during early stages of cell evolution.

The genomes of heat-loving prokaryotes

Comparison of archaeal and bacterial genomes has led to the identification of a number of thermophilic/hyperthermophilic-specific signatures, including the low abundance of the dinucleotide CG in their DNA [21], amino acid compositional biases [48], reduced protein-encoding gene length [55, 48], and the presence of reverse gyrase, an ATP-dependent topoisomerase described as a hallmark of a heat-loving lifestyle [13]. As discussed below, hyperthermophilic genomes have additional characteristic traits.
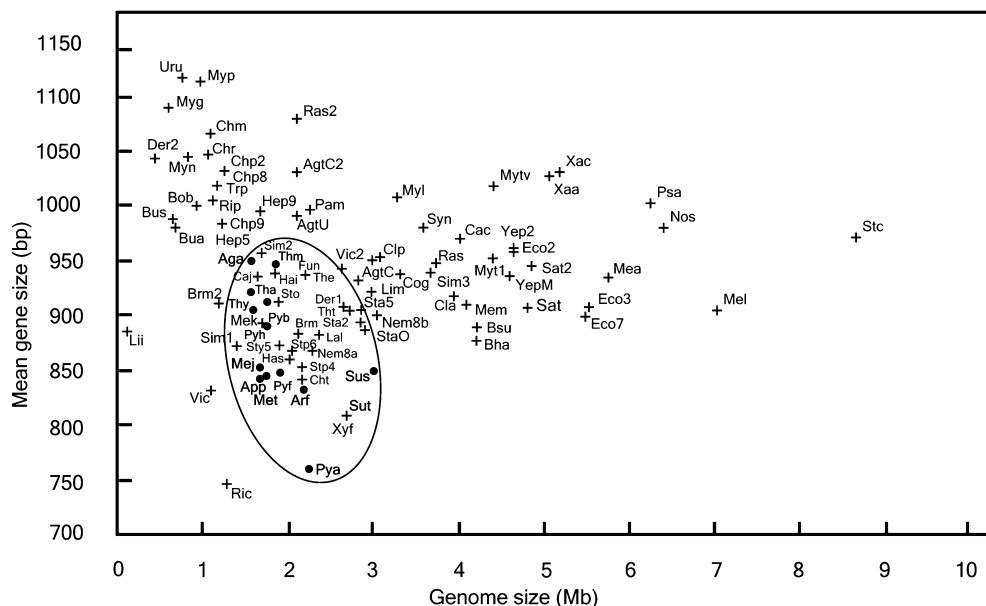
Although considerable variations in DNA content exist within closely related bacterial species and strains, the available data suggest that genome sizes of each of the three domains appear to lie within defined ranges [40]. As part of an attempt to study the size and organization of prokaryotic chromosomes, a database was constructed with 641 archaeal and bacterial genome sizes determined by pulsed-field gel electrophoresis (PFGE), published as of December 2002. This database is available upon request, and will be published elsewhere.

Figure 1 shows the genome size distribution of our sample. Mesophilic proteobacteria have the largest genomes, ranging from 0.448 Mb for *Buchnera* sp. to 9.7 Mb for *Azospirillim lipoferum* 59B. The smallest bacterial genome sizes correspond to obligate symbionts such as *Buchnera* sp. and pathogens such as *Mycoplasma*, *Chlamydia* and *Rickettsia*, whose small DNA content is a derived character that reflects secondary gene loss due to their parasitic lifestyles.

Of course, the data summarized in Fig. 1 are biased by an inadequate sampling that does not fully represent

Mean gene size (bp) — axis values: 1150, 1100, 1050, 1000, 950, 900, 850, 800, 750, 700

Genome size (Mb) — axis values: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

the true levels of microbial biodiversity, but is clearly skewed, on the one hand, towards pathogenic bacteria and, on the other, to extremophilic archaea. Nevertheless, it provides useful insights into the size and organization of prokaryotic genomes. Although thermophilic and hyperthermophilic bacterial and archaeal genomes follow a trend similar to that of their mesophilic counterparts, they depart from a normal distribution and fall within a well-defined size range (from 0.5 Mb for the thermophilic ectosymbiont *Nanoarchaeon equitans,* to the 5.10 Mb of the facultative thermophilic *Methanosarcina acetivorans*), with a maximum at approximately 2 Mb. However, this size range does not necessarily reflect a correlation between DNA content, heat-loving microbial lifestyles and antiquity, since many different mesophilic bacterial species, including *Leptospira,* green-sulfur bacteria, cyanobacteria, spirochaetes, fusobacteria and actinobacteria, are endowed with similarly small-sized chromosomes.

Genomic analysis has shown that thermophilic and hyperthermophilic genomes are endowed with smaller protein-encoding genes than their mesophilic counterparts [48, 55]. Detailed statistical analysis of 56 complete genomes, including seven eukaryotes, 14 archaeal and 35 bacterial species, has shown that the mean protein length of heat-loving prokaryotes (283 ± 5.8) is significantly smaller than in mesophiles (340 ± 9.4) [48]. It is possible that these reduced gene sizes are correlated with an extremophilic lifestyle, such as protein thermostability. As shown in Fig. 2, however, reduced gene size is a polyphyletic trait: small protein-encoding genes that fall within the same size range of hyperthermophilic genes are also found in a wide diversity of non-extremophiles, including proteobacteria, green-sulfur bacteria, low GC gram-positives, fusobacteria, and mesophilic euryarchaea.

Like their mesophilic counterparts, hyperthermophilic genes are endowed with simple sequences, i.e. homopolymeric tracts and tandem arrays of multiple short repeat motifs. These low-complexity regions have their origin in mutational processes, such as slipped-strand mispairing and unequal crossing-over, that take place during DNA replication and are known to represent a major source of genetic variation in pathogenic prokaryotes [31]. Analysis of all the completely sequenced hyperthermophilic and thermophilic genomes available as of December 2002 shows that the natural amino acid composition of each proteome is enhanced with respect to its corresponding simple sequences, which have a compositional bias as shown by the abundance of small, α-helix forming amino acids, i.e., alanine, leucine, lysine, serine and glutamic acid (Becerra, Cocho, Delaye and Lazcano, unpublished results). As shown in Fig. 3, however, simple sequences in hyperthermophiles are clearly enriched in glutamic acid. The stability of the α-helix structure of glutamic acid homopolymers under acid pH values [35] probably explains why, with the exception of *Thermoplasma acidophilum*, simple sequences of acid-resistant heat-loving prokaryotes tend to be rich in this amino acid. Enrichment of glutamic acid in extremophilic simple sequences explains the relative abundance in hyperthermophilic genomes, as noted by Tekaia et al. [48].

The faulty records of archaean life

As shown by recent debates, the identification of the oldest paleontological traces of life can be a highly contentious issue. The early archaean geological record is scarce, and most of the preserved rocks have been metamorphosed to a considerable extent. However, the evidence suggests that life emerged on Earth as soon as it was possible to do so. Although the biological origin of the microstructures interpreted as cyanobacterial remnants in the 3.5×10^9 year old Apex sediments of the
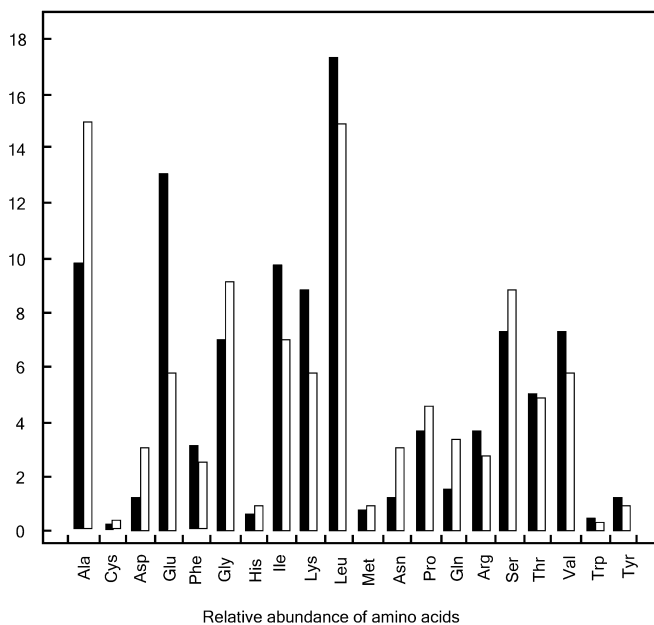
Fig. 3 Relative abundances of amino acids in simple sequences in all available proteomes as of December, 2002. Simple sequences were identified using the SEG program [54], which identifies low-complexity regions in which an enhanced concentration of short repeats not due to chance events can be detected. *White bars* show the average simple-sequence amino acid composition of mesophiles, and *dark bars* show those of hyperthermophilic prokaryotes (80 °C or more). The hyperthermophilic species represented here are *Pyrococcus furiosus*, *P. horikopshi*, *P. abyssi*, *Aeropyrum pernix*, *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Sulfolobus solfataricus* and *S. tokodaii*

Australian Warrawoona formation [38] have been questioned [3], there is additional paleontological evidence that highly diverse microbial communities were thriving during the early and middle Archaean [32, 49].

Unfortunately, it is unlikely that data on how life originated will be provided by the geological record. There is no direct evidence of the environmental conditions on the Earth at the time of the origin of life, nor is there any fossil register of the evolutionary processes that preceded the appearance of the first cells. Direct information is lacking not only on the composition of the terrestrial atmosphere during the period of the origin of life, but also on the temperature, ocean pH values, and other general and local environmental conditions that may or may not have been important for the emergence of living systems.

The attributes of the first living organisms are also unknown. They were probably simpler than any cell now alive and may have lacked not only protein-based catalysis, but perhaps even the familiar genetic macromolecules, with their sugar-phosphate backbones. It is possible that the only property they shared with extant organisms was the structural complementarity between monomeric subunits of replicative informational polymers, e.g. the joining together of residues in a growing chain whose sequence is directed by preformed polymers. Such ancestral polymers may have not even involved

nucleotides. Accordingly, 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 prokaryotes represented in molecular phylogenies.

The rooting of universal cladistic trees determines the directionality of evolutionary change and allows ancestral characters to be distinguished from those that were derived. Determination of the rooting point of a tree normally imparts polarity to most or all characters. It is, however, important to distinguish between ancient and primitive organisms. Organisms located near the root of universal rRNA-based trees are cladistically ancient, but they are not endowed with a primitive molecular genetic apparatus, nor do they appear to be more rudimentary in their metabolic abilities than their aerobic counterparts. Primitive living systems would initially refer to pre-RNA worlds, in which life may have been based on polymers using backbones other than ribose-phosphate and possibly bases different from adenine, uracil, guanine and cytosine, followed by a stage in which life was based on RNA as both genetic material and catalysts [23].

Molecular cladistics may provide clues to some very early stages of biological evolution, but 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, i.e., an RNA/protein world. Older stages are not yet amenable to molecular phylogenetic analysis. A cladistic approach to the origin of life itself is not feasible, since all possible intermediates that may have once existed have long since vanished.

Was the last common ancestor a hyperthermophile?

The variations of traits common to extant species can be easily explained as the outcome of divergent processes from an ancestral life form that existed prior to the separation of the three major biological domains, i.e., the last common ancestor (LCA) or cenancestor. No paleontological remains will bear testimony of its existence, as the search for a fossil of the cenancestor is bound to prove fruitless. From a cladistic viewpoint, the LCA is merely an inferred inventory of features shared among extant organisms, all of which are located at the tip of the branches of molecular phylogenies. However, if the term "universal distribution" is restricted to its most obvious sense, i.e., that of traits found in all completely sequenced genomes now available, then quite unexpectedly the resulting repertoire is formed by relatively few features and by incompletely represented biochemical processes [8]. Surprisingly, some of the most likely a priori candidates for strict universality, such as those sequences involved in DNA replication, have also turned out to be poorly preserved [11].

Analysis of an increasingly large number of completely sequenced cellular genomes has revealed major discrepancies in the topology of rRNA trees. Very often

these differences have been interpreted as evidence of horizontal gene-transfer events between different species, questioning the feasibility of the reconstruction and proper understanding of early biological history [10]. There is clear evidence that genomes have a mosaic-like nature whose components come from a wide variety of sources. Depending on their different advocates, a wide spectrum of mix-and-match recombination processes have been described, ranging from the lateral transfer of few genes via conjugation, transduction or transformation, to cell fusion events involving organisms from different domains.

The resulting reticulate phylogenies greatly complicate the inference of cenancestral traits. Driven in part by the impact of lateral gene acquisition, as revealed by the discrepancies of different gene phylogenies with the canonical rRNA tree, and in part by the surprising complexity of the universal ancestor, as suggested by direct backtrack characterizations of the oldest node of universal cladograms, Woese [53] has argued that the LCA was not a single organismic entity, but rather a highly diverse population of metabolically complementary, cellular progenotes endowed with multiple, small linear chromosome-like genomes that benefited from massive multidirectional horizontal transfer events. According to this viewpoint, the development of the essential features of translation and of metabolic pathways took place before the earliest branching event, but what led to the three domains was not a single ancestral lineage, rather a rapidly differentiating community of genetic entities. This communal ancestor occupied as a whole the node located at the bottom of the universal tree, in which decreasing sequence exchange and increasing genetic isolation would eventually lead to the observed tripartite division of the biosphere.

Did the hypothetical communal progenote ancestor proposed by Woese [53] diverge sharply into the three domains soon after the appearance of the code and the establishment of translation? Not necessarily, since inventories of LCA genes clearly include sequences that originated in different pre-cenancestral epochs. The origin of the mutant sequences ancestral to those found in all extant species, and the divergence of the Bacteria, Archaea, and Eukarya were not synchronous events, i.e., the separation of the primary domains took place later, perhaps even much later, than the appearance of the genetic components of their LCA [8].

Universal gene-based phylogenies ultimately reach a single universal entity, but the bacterial-like LCA [8] that we favor was not alone. Company must have been provided by its siblings, a population of entities similar to it that existed throughout the same period. They may not have survived, but some of their genes did if they became integrated via lateral transfer into the LCA genome. The cenancestor should thus be considered as the evolutionary outcome of a series of ancestral events, including lateral gene transfer, gene losses, and paralogous duplications, that took place before the separation of Bacteria, Archaea, and Eukarya.

Comparisons of combined ortholog protein data sets that exclude sequences which may have undergone lateral transfer are consistent with rRNA trees [5]. Genomic trees also exhibit an excellent broad-level agreement with rRNA-based phylogenies [47]. Genomic trees are not cladograms but phenograms, i.e., they are hierarchical representations of similarities and differences in gene content, in which the presence or absence of a sequence is counted as a character. Since different lineages evolve at different rates, such overall similarity may be an equivocal indicator of genealogical relationships. Nevertheless, these trees are rooted in the same area as rRNA phylogenies, which suggests that massive lateral transfer events between distant groups has not obliterated the early history of life. Thus, although hyperthermophiles may be displaced from their basal position if molecular markers other than elongation factors or ATPase subunits are employed, or if alternative phylogeny-building methodologies are used [4], it can still be argued that rRNA-based phylogenies provide one of the best-preserved historical records of cell evolution [53].

The recognition that the deepest branches in rooted universal phylogenies are occupied by hyperthermophiles does not by itself provide conclusive proof of a heat-loving LCA. Analysis of the correlation of the optimal growth temperature of prokaryotes and the $G + C$ nucleotide content of 40 rRNA sequences through a complex Markov model has led Galtier et al. [16] to conclude that the universal ancestor was a mesophile. This possibility has been contested by Di Giulio [9], who has argued for a thermophilic or hyperthermophilic LCA. However, since the time factor is absent from the methodology developed by Galtier et al. [16], the inferred low $G + C$ content of the cenancestral rRNA does not necessarily belong to the cenancestor itself, but may correspond to a mesophilic predecessor that may have been located along the trunk of the universal tree.

Chemical evolution and extreme environments

The hypothesis that the first organisms were anaerobic heterotrophs is based on the assumption that abiotic organic compounds were a necessary precursor for the appearance of life. The first successful synthesis of organic compounds under plausible primordial conditions was accomplished 50 years ago by the action of electric discharges acting for a week over a mixture of CH_4, NH_3, H_2, and H_2O, and led to complex mixture of monomers that included racemic mixtures of several proteinic amino acids, in addition to hydroxy acids, urea and other molecules [27]. Prebiotic synthesis of amino acids largely proceeds by a Strecker synthesis that involves the aqueous-phase reactions of highly reactive intermediates

$$RCHO + HCN + 2\,NH_3 \rightleftharpoons RCH(NH_2)CN + H_2O$$

$$RCH(NH_2)CN + 2\,H_2O \longrightarrow RCH(NH_2)COOH + NH_3$$
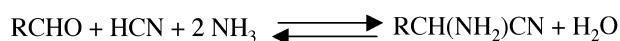
Detailed studies of the equilibrium and rate constants of these reactions demonstrated that both amino acids and hydroxy acids can be synthesized at high dilutions of HCN and aldehydes in a simulated primitive ocean. The reaction rates depend on temperature, pH, HCN, NH_3, and aldehyde concentrations, and are rapid on a geological time scale; the half-lives for the hydrolysis of the intermediate products in the reactions, amino nitriles and hydroxy nitriles, are less than a 1,000 years at 0 °C, and there are no known slow steps [30].

The remarkable ease by which adenine can be synthesized by the aqueous polymerization of ammonium cyanide demonstrated the significance of HCN and its derivatives in prebiotic chemistry [33]. As summarized elsewhere [30], the prebiotic importance of HCN has been further substantiated by the discovery that the hydrolytic products of its polymers include amino acids, purines, and pyrimidines. The reaction of cyanoacetylene or cyanoacetaldehyde (a hydrolytic derivative of HCN) with urea leads to high yields of cytosine and uracil, especially under simulated evaporating pond conditions which increase the urea concentration [36].

The ease of formation under reducing conditions ($CH_4 + N_2$, $NH_3 + H_2O$, or $CO_2 + H_2 + N_2$) of amino acids, purines, and pyrimidines in one-pot reactions strongly suggests that these molecules were present in the prebiotic broth. In addition, experimental evidence suggests that urea, alcohols, sugars formed by the non-enzymatic condensation of formaldehyde, a wide variety of aliphatic and aromatic hydrocarbons, urea, carboxylic acids, and branched and straight fatty acids, including some which are membrane-forming compounds, were also components of the primitive soup. The remarkable coincidence between the molecular constituents of living organisms and those synthesized in prebiotic experiments is too striking to be fortuitous, and the robustness of this type of chemistry is supported by the occurrence of most of these biochemical compounds in the 4.5×10^9-year-old Murchison carbonaceous meteorite, which also yielded evidence of liquid water in its parent body [12].

A major advantage of high temperatures is that chemical reactions go faster, and the primitive enzymes, once they appeared, could have thus been less efficient but nonetheless effective. However, the price paid is manifold: high-temperature regimes would lead to: (a) reduced concentrations of volatile intermediates, such as HCN, H_2CO and NH_3; (b) lower steady-state concentrations of prebiotic precursors like HCN, which at temperatures a little above 100 °C undergoes hydrolysis to formamide and formic acid and, in the presence of ammonia, to NH_4HCO_2 (Structure 3).

$$HCN \xrightarrow{H_2O} H-\overset{\overset{O}{\|}}{C}-NH_2 \xrightarrow{H_2O} H-\overset{\overset{O}{\|}}{C}-OH \xrightarrow{NH_4^+} NH_4HCO_2$$
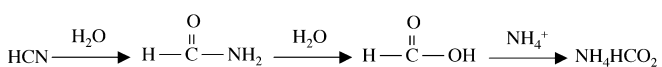
(c) instability of reactive chemical intermediates like amino nitriles ($RCHO(NH_2)CN$), which play a central role in the Strecker synthesis of amino acids (see Structure 1); and (d) loss of organic compounds by thermal decomposition and diminished stability of genetic polymers.

Extremophilic genomes are protected against thermal decomposition by a number of enzyme-dependent mechanisms [18], but these would have not been available during prebiotic times or at the time of the origin of life. In fact, the existence of an RNA world with ribose appears to be incompatible with a (hyper)thermophilic environment [29]. Survival of nucleic acids is limited by the hydrolysis of phosphodiester bonds [24], and the stability of Watson-Crick helices (or their pre-RNA equivalents) is strongly diminished by high-temperatures. For an RNA-based biosphere, reduced thermal stability on the geologic time scale of ribose and other sugars is the worst problem, but the situation is equally bad for pyrimidines, purines and some amino acids. As summarized elsewhere [23, 39], measurements by different groups have shown that the half-life of ribose at 100 °C and pH 7 is only 73 min, and other sugars (2-deoxyribose, ribose 5-phosphate, and ribose 2,4-biphosphate) have comparable half-lives. The half-life for hydrolytic deamination of cytosine at 100 °C is 19–21 days, although at 100 °C the half-life of uracil is approximately 12 years. At 100 °C, the thermal stability of purines is also reduced: 204–365 days for adenine, with guanine having a low half-life.

A hyperthermophilic pyrite-dependent origin of life?

An alternative to the problem of low half-lives of biochemical monomers at temperatures of 100 °C or more is to assume an autotrophic origin of life. Such proposals are periodically resurrected, but they are generally made without supportive evidence. The most elaborate chemoautotrophic-origin-of-life scheme has been proposed by Wächtershäuser [50]. According to this hypothesis, life began with the appearance of an autocatalytic two-dimensional chemolithotrophic metabolic system based on the formation of the highly insoluble mineral pyrite. The synthesis in activated form of organic compounds such as amino acid derivatives, thioesters and keto acids is assumed to have taken place on the surface of FeS and FeS_2 in environments that resembled those of deep-sea hydrothermal vents. Replication followed the appearance of non-organismal iron-sulfide-based two-dimensional life, in which chemoautotrophic carbon fixation took place by a reductive citric acid cycle, or reverse Krebs cycle, of the type originally described for the photosynthetic green sulfur bacterium *Chlorobium limicola*. Molecular phylogenetic trees show that this mode of carbon fixation and its modifications (such as the reductive acetyl-CoA or the reductive malonyl-CoA pathways) are found in anaerobic archaea and the most deeply divergent eubacteria, which has been interpreted as evidence of its primitive character [25].

The reaction $FeS + H_2S \rightarrow FeS_2 + H_2$ is a very favorable one. It has an irreversible, highly exergonic character with a standard free-energy change $\Delta G^0 = -9.23$ kcal/mol, which corresponds to a reduction

potential $E^0 = -620$ mV. Thus, the FeS/H$_2$S combination is a strong reducing agent, and has been shown to provide an efficient source of electrons for the reduction of organic compounds under mild conditions. Pyrite-mediated CO$_2$ reduction to amino acids, purines and pyrimidines is yet to be achieved. However, as reviewed elsewhere [6, 20, 25], the FeS/H$_2$S combination has been shown to: (a) reduce nitrate and acetylene; (b) induce peptide-bond formation that results from the activation of amino acids with carbon monoxide and (Ni, Fe)S; and (c) to induce the synthesis of acetic acid and pyruvic acid from CO under simulated hydrothermal conditions in the presence of sulfide minerals [6, 20, 25]. However, support for Wächtershäuser's central tenets is meager. Life does not consist solely of metabolic cycles, and none of these experiments prove that enzymes and nucleic acids are the evolutionary outcome of multistep autocatalytic metabolic cycles surface-bounded to FeS/FeS$_2$ or some other mineral. In fact, experiments using the FeS/H$_2$S combination are also compatible with a more general, modified model of the primitive soup in which pyrite formation is recognized as an important source of electrons for the reduction of organic compounds [2].

Summary and conclusions

As the initially molten young Earth cooled down, global temperatures of 100 °C must have been reached but could not have persisted for more than 20 million years [42]. Deep-sea hydrothermal vents and other local high-temperature milieus have existed throughout the history of the planet and have played a major role in shaping the early environments. However, the rates of thermal decomposition of amino acids, nucleobases, and genetic polymers are very short on the geological time scale and argue against a hot origin of life in such extreme environments.

Since high salt concentrations protect DNA and RNA against heat-induced damage [26, 46], this and other non-biological mechanisms, such as adsorption to minerals surfaces and formation of clay–nucleic acid complexes [15] might have played a significant role in the preservation of organic compounds and genetic polymers in the primitive environments. However, such mechanisms would be inefficient at temperatures above 100 °C. Because adsorption involves the formation of weak noncovalent bonds, mineral-based concentration and protection would have been most effective at low temperatures [43]; at high temperatures any adsorbed monomers would drift away into the surrounding aqueous environment and become hydrolyzed. However, some minerals could also have the opposite effect: as shown by the Cu^{+2}-montmorillonite catalyzed decomposition of adenine to hypoxanthine [45], the association of organic compounds with some minerals may in fact reduce their half-lives.

If hyperthermophily is not truly primordial, then heat-loving lifestyles may be relics of a secondary adaptation that evolved after the origin of life and before or soon after the separation of the major lineages. As argued here, the so-called root of universal trees does not correspond to the first living system, but is the tip of a trunk of still undetermined length in which the history of a long (but not necessarily slow) series of archaic evolutionary events such as an explosion of gene families and multiple events of lateral gene transfer are still preserved. Is it possible that traces of the emergence of hyperthermophily persist in the molecular records of earliest biological evolution somewhere along the trunk of rRNA-based phylogenic trees? If hyperthermophiles were not the first organisms, then their basal position in molecular trees could be explained as: (a) a relic from early archean high-temperature regimes that may have resulted from a severe impact regime [17, 41]; (b) adaptation of Bacteria to extreme environments by lateral transfer of reverse gyrase [14] and other thermo-adaptative traits from heat-loving Archaea; and (c) outcompetition of older mesophiles by hyperthermophiles originally adapted to stress-inducing conditions other than high temperatures [29].

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. Like vegetation in a mangrove, the roots of universal phylogenetic trees are submerged in the muddy waters of the prebiotic broth, but how the transition from the non-living to the living took place is still unknown. Given the huge gap existing in current descriptions of the evolutionary transition between the prebiotic synthesis of biochemical compounds and the LCA of all extant living beings, it is probably naive to attempt to describe the origin of life and the nature of the first living systems from molecular phylogenies. A high-temperature origin of life may be possible, but if this was the case then it could have not involved the usual purines and pyrimidines, or other biochemical monomers.

References

1. Achenbach-Richter L, Gupta R, Stetter KO, Woese CR (1987) Were the original eubacteria thermophiles? System Appl Microbiol 9:34–39
2. Bada JL, Lazcano A (2002) Some like it hot, but not biomolecules. Science 296:1982–1983
3. Brasier MD, Green OR, Jephcoat AP, Kleppe AK, van Kranendonk MJ, Lindsay JF, Steele A, Grassineau NV (2002) Questioning the evidence for Earth's oldest fossils. Nature 416:76–81
4. Brochier C, Philippe H (2002) A non-hyperthermophilic ancestor for Bacteria. Nature 417:244

5. Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ (2001) Universal trees based on large combined protein sequence data sets. Nat Genet 28:281–285

6. Cody GD, Boctor NZ, Filley TR, Hazen RM, Scott JH, Sharma A, Yoder HS Jr. (2000) Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate. Science 289:1337–1340

7. Corliss JB, Baross JA, Hoffman SE (1981) An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. Oceanologica Acta (Suppl.) 4:59–69

8. Delaye L, Becerra A, Lazcano A (2002) The nature of the last common ancestor. In: Ribas de Pouplana L (ed) The genetic code and the origin of life. Landes Bioscience, Georgetown (in press)

9. Di Giulio M (2000) The universal ancestor lived in a thermophilic or hyperthermophilic environment. J Theor Biol 203:203–213

10. Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2129

11. Edgell RD, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. Cell 89:995–998

12. Ehrenfreund P, Irvine W, Becker L, Blank J, Brucato J, Colangeli L, Derenne S, Despois D, Dutrey A, Fraaije H, Lazcano A, Owen T, Robert F (2002) Astrophysical and astrochemical insights into the origin of life. Reports Prog Phys 65:1427–1487

13. Forterre P (2002) A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. Trends Genet 18:236–237

14. Forterre P, Bouthier de la Tour C, Philippe H, Duguet M (2000) Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from Archaea to Bacteria. Trends Genet 16:152–154

15. Franchi M, Bramanti E, Morassi Bonzi LM, Orioli PL, Vettori C, Gallori E (1999) Clay-nucleic acid complexes: characteristics and implications for the preservation of genetic material in primeval habitats. Origins Life Evol Biosph 29:297–315

16. Galtier N, Tourasse N, Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. Science 283:220–221

17. Gogarten-Boekels M, Hilario E, Gogarten JP (1994) The effects of heavy meteorite bombardment on the early evolution of life—a new look at the molecular record. Origins Life Evol Biosph 25:78–83

18. Grogan DW (1998) Hyperthermophiles and the problem of DNA instability. Mol Microbiol 28:1043–1049

19. Holm NG (ed) (1992) Marine hydrothermal systems and the origin of life. Kluwer Academic, Dordrecht

20. Huber C, Wächtershäuser G (1998) Peptides by activation of amino acids with CO on (Ni, Fe)S surfaces: implications for the origin of life. Science 281:670–672

21. Karlin S, Mrázek J (1998) Prokaryotic genome-wide comparisons and evolutionary implications. In: de Bruijn FJ, Lupski JR, Weinstock GM (eds) Bacterial Genomes: physical structure and analysis. Kluwer Academic, Boston, pp 196–212

22. Lamarck JB (1804) Zoological Philosophy: an exposition with regard to the Natural History of Animals. The University of Chicago Press, Chicago (Translated 1984), 458 pp

23. Levy M, Miller SL (1998) The stability of the RNA bases: implications for the origin of life. Proc Natl Acad Sci USA 95:7933–7938

24. Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715

25. Maden BEH (1995) No soup for starters? Autotrophy and the origins of metabolism. Trends Biochem Sci 20:337–341

26. Marguet E, Forterre P (1994) DNA stability at temperatures typical for hyperthermophiles. Nucleic Acid Res 22:1681–1686

27. Miller SL (1953) A production of amino acids under possible primitive Earth conditions. Science 117:528

28. Miller SL, Bada JL (1988) Submarine hot springs and the origin of life. Nature 334:609–611

29. Miller SL, Lazcano A (1995) The origin of life—did it occur at high temperatures? J Mol Evol 41:689–692

30. Miller SL, Lazcano A (2002) Formation of the building blocks of life. In: Schopf JW (ed) Life's origin: the beginnings of biological evolution. California University Press, Berkeley, pp 78–112

31. Moxon ER (1999) Whole-genome analysis of pathogens. In: Stearns SC (ed) Evolution in health and disease. Oxford University Press, New York pp 191–204

32. Nisbet EG, Sleep NH (2001) The habitat and nature of early life. Nature 409:1083–1091

33. Oró J (1960) Synthesis of adenine from ammonium cyanide. Biochem Biophys Res Commun 2:407–412

34. Pace NR (1991) Origin of life—facing up to the physical setting. Cell 65:531–533

35. Palmer T (1995) Understanding enzymes. Prentice Hall, Hertfordshire

36. Robertson MP, Miller SL (1995) An efficient prebiotic synthesis of cytosine and uracil. Nature 375:772–774

37. Schopf JW (ed) (1983) Earth's earliest biosphere: its origin and evolution. Princeton University Press, Princeton

38. Schopf JW (1993) Microfossils of the early Archaean Apex chert: new evidence of the antiquity of life. Science 260:640–646

39. Shapiro R (1995) The prebiotic role of adenine: a critical analysis. Origins Life Evol Biosph 25:83–98

40. Shimkets LJ (1998) Structure and sizes of the genomes of the Archaea and Bacteria. In: de Bruijn FJ, Lupski JR, Weinstock GM (eds) Bacterial genomes: physical structure and analysis. Kluwer Academic, Boston, pp 5–11

41. Sleep NH, Zahnle KJ, Kastings JF, Morowitz HJ (1989) Annihilation of ecosystems by large asteroid impacts on the early Earth. Nature 342:139–142

42. Sleep NH, Zahnle K, Neuhoff PS (2001) Initiation of clement surface conditions on the earliest Earth. Proc Natl Acad Sci USA 98:3666–3672

43. Sowerby SJ, Mörth C-M, Holm NG (2001) Effect of temperature on the adsorption of adenine. Astrobiology 1:481–487

44. Stetter KO (1994) The lesson of archaebacteria. In: Bengtson S (ed) Early life on earth: Nobel symposium no. 84. Columbia University Press, New York, pp 114–122

45. Strasak M, Sersen F (1991) An unusual reaction of adenine and adenosine on montmorillonite: a new way of prebiotic synthesis of some purine nucleotides? Naturwissenschaften 78:121–122

46. Tehei M, Franzetti B, Maurel M-C, Vergne J, Hountondji C, Zaccai G (2002) The search for traces of life: the protective effect of salt on biological macromolecules. Extremophiles 6:427–430

47. Tekaia F, Lazcano A, Dujon B (1999) The genomic tree as revealed from whole proteome comparisons. Genome Res 9:550–557

48. Tekaia F, Yeramian E, Dujon B (2002) Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. Gene 297:51–60

49. Van Kranendonk MJ (2002) The flourishing of early life on Earth at hydrothermal vents: geological evidence from the 3.49–3.43 Ga Warrawoona Group, Pilbara Craton, Western Australia. Abstracts of the IAU Symposium 213 Bioastronomy 2002: Life among the stars. Australian Centre for Astrobiology, Hamilton Island, Great Barrier Reef, Australia, July 8–12, 2002, p 33

50. Wächtershäuser G (1988) Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52:452–484

51. White RH (1984) Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250 °C. Nature 310:430–432

52. Wilde SA, Valley JW, Peck WH, Graham CM (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. Nature 409:175–178

53. Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99:8742–8747

54. Wootton J, Federhen S (1993) Statistics of local complexity in amino acid sequence and sequence database. Compt Chem 17:149–163

55. Zhang J (2000) Protein-length distributions for the three domains of life. Trends Genet 16:107–109