THE EFFECTS OF HEAVY METEORITE BOMBARDMENT ON THE EARLY EVOLUTION – THE EMERGENCE OF THE THREE DOMAINS OF LIFE

MARIA GOGARTEN-BOEKELS, ELENA HILARIO and J. PETER GOGARTEN
Dept. Molecular and Cell Biology, University of Connecticut, 75 North Eagleville Road, Storrs, CT 06269-3044, U.S.A.

(Received November 1, 1993)

Abstract. A characteristic of many molecular phylogenies is that the three domains of life (Bacteria, Archaea, Eucarya) are clearly separated from each other. The analyses of ancient duplicated genes suggest that the last common ancestor of all presently known life forms already had been a sophisticated cellular prokaryote. These findings are in conflict with theories that have been proposed to explain the absence of deep branching lineages. In this paper we propose an alternative scenario, namely, a large meteorite impact that wiped out almost all life forms present on the early Earth. Following this nearly complete frustation of life on Earth, two surviving extreme thermophilic species gave rise to the now existing major groups of living organisms, the Bacteria and Archaea. [The latter also contributed the major portion to the nucleo-cytoplasmic component of the Eucarya]. An exact calibration of the molecular record with regard to time is not yet possible. The emergence of Eucarya in fossil and molecular records suggests that the proposed late impact should have occurred before 2100 million years before present (BP). If the 3500 million year old microfossils [Schopf, J. W. 1993: Science 260: 640–646] are interpreted as representatives of present day existing groups of bacteria (*i.e.*, as cyanobacteria), then the impact is dated to around 3700 million years BP.

The analysis of molecular sequences suggests that the separation between the Eucarya and the two prokaryotic domains is less deep then the separation between Bacteria and Archaea. The fundamental cell biological differences between Archaea and Eucarya were obtained over a comparatively short evolutionary distance (as measured in number of substitution events in biological macromolecules).

Our interpretation of the molecular record suggests that life emerged early in Earth's history even before the time of the heavy bombardment was over. Early life forms already had colonized extreme habitats which allowed at least two prokaryotic species to survive a late nearly ocean boiling impact. The distribution of ecotypes on the rooted universal tree of life should not be interpreted as evidence that life originated in extremely hot environments.

1. Introduction

Two main sources are available to study the early evolution of life on this planet: the fossil record and the molecular, biochemical and anatomical records that survived as a heritage in extant organisms. The early Precambrian fossil record is sparse; however, significant progress has been made in identifying and characterizing microfossils from this era (Schopf, 1992). Microfossils of prokaryotes date back to at least 3500 million years before present (BP). The filamentous morphology and size of these ancient microfossils and their occurence in stromatolite-like structures were interpreted to suggest that these microfossils represent photosynthetic oxygen producing bacteria (Awramik, 1992; Schopf, 1992, 1993).

The molecular record is stored in biological macromolecules (e.g., Schwartz and Dayhoff, 1978) and in biochemical pathways (e.g., Granick, 1957; Wächtershäuser, 1990) of extant organisms. During the past decade successful attempts were made

Origins of Life and Evolution of the Biosphere **25:** 251–264, 1995.

to directly extend the molecular record to fossils (Pääbo, 1993). However, the recovery of genetic information directly from fossilized organisms is limited to comparatively young specimen; currently, the oldest are from 40 million years BP. The vulnerability of information carrying macromolecules (DNA, RNA, protein) leaves little hope to expand the availability of molecular fossils into the Precambrian era.

The comparison of DNA, RNA and protein sequences from different extant species allows one to reconstruct the evolutionary history of these molecules. If their evolution reflects speciation, these molecules can also be used as markers for the organismal evolution. Different molecules contain information that is useful at different phylogenetic levels (cf. Bruns *et al.*, 1991).

Sequences that are under a high selection pressure have a low substitution rate. These sequences can be used to study the early evolution of life. A perfect molecular clock should experience substitutions with a constant rate. Ochman and Wilson (1987) postulated a nearly clock like behavior for some molecular marker molecules during the last 2000 million years; however, changes in the substitution rates by up to a factor of 13 have been observed (Ohta, 1991).

Currently, a satisfactory calibration of molecular phylogenies is only possible for the more recent evolution (Moran *et al.*, 1993). The assumption of constant substitution rates for the early evolution is hardly justified; but in the absence of alternatives this is often the only available avenue (e.g., Eigen *et al.*, 1989; Sogin, 1992). The calibration of molecular phylogenies using microfossils is possible (see below); however, the uncertainties and error margins are large (compare Ochman and Wilson, 1987), therefore, the calibration of ancient molecular phylogenies with respect to time remains ambiguous.

The interpretation of biological records (fossil and molecular) is aided by geological and astrophysical findings and theories. As discussed by Miller (1992) the conditions on the early Earth were very different from present day environments. In particular, the analysis of the moon cratering record suggests that large impacts were much more frequent than today (Mahler and Stevenson, 1988). Overbeck and Fogleman (1990) concluded from astrophysical considerations that life which existed at around 3800 million years BP would very likely have been destroyed by giant impacts.

In this paper we discuss the molecular records in light of both, the fossil evidence and the cratering record. We evaluate different scenarios for the early evolution of life, taking into account the recent findings of early duplicated genes and the inferred properties of the last common ancestor.

The Molecular Record

The Three Domains and the Progenote

From the study of 16S ribosomal RNA it became apparent that the prokaryotes can be divided into two groups (Woese and Fox, 1977b). Originally these were named

the eubacterial and the archaebacterial Ur-Kingdoms. More recently these two Ur-Kingdoms were renamed into the domains Bacteria and Archaea (Woese *et al.*, 1990). The division of the prokaryotes is supported by other molecular phylogenies (e.g., ATPases, elongation factors, RNA polymerases; Gogarten *et al.*, 1989; Iwabe *et al.*, 1989; Puhler *et al.*, 1989) and by many other biochemical characters (see Zillig *et al.*, 1992, for a recent summary). The third domain of life is constituted by the eukaryotes or Eucarya. When the three domains were first recognized it was not obvious which of these groups were more closely related to each other.

Woese and Fo//x (1977a) introduced the term 'progenotic stage' to denote a primitive stage of development that existed before a defined relation between genotype and phenotype of an organism had been established, i.e., before the prokaryotic stage. Woese and Fox (1977a) suggested that '*It is at this progenotic state, not the procaryote stage, that the line of descent leading to the eucaryotic cytoplasm diverged from the bacterial lines of descent.*' Certainly, early life forms must have existed whose organizational level corresponds to the progenotic stage; however, as will be discussed below, the analysis of duplicated genes and the consideration of the many shared characteristics of extant cellular life suggest that the last common ancestor already had reached the prokaryotic stage, i.e., the last common ancestor was not a progenote as originally defined, but a prokaryote not too dissimilar from extant prokaryotes (Gogarten and Taiz, 1992; Lazcano, 1993a).

THE ORIGINS OF THE EUKARYOTES

The endosymbiont theory (cf. Margulis, 1981) maintains that some eucaryal cell organelles evolved from Bacteria that functioned as endosymbionts within a host cell. In the case of mitochondria and plastids the bacterial origin has been verified by studies of molecular evolution (e.g., Schwartz and Dayhoff, 1978). The information provided by ancient duplicated genes (ATPase subunits, dehydrogenases, elongation factors, ^{met}tRNAs; Gogarten *et al.*, 1989; Iwabe *et al.*, 1989) revealed that a major portion of the nucleocytoplasm of the host cell evolved from an Archaeum-like prokaryote. This finding is further corroborated by the similarities in the transcription machinery in Eucarya and Archaea (Puhler *et al.*, 1989). Other molecular markers appear to contradict the close association between Archaea and the eucaryal nucleocytoplasm (glutamate dehydrogenase, Benachenhou-Lafha *et al.*, 1993; carbamylphosphate synthetase, Lazcano, Puente, Gogarten, unpublished; glutamine synthase, Kumada *et al.*, 1993; Tiboni *et al.*, 1993; heat shock proteins, Gupta and Golding, 1993). More detailed analyses are necessary to decide whether these molecular phylogenies represent cases of horizontal gene transfer, additional major contributions to the eukaryotic nucleocytoplasm, or ill resolved molecular phylogenies (see Hilario and Gogarten, 1993, for further discussion).

TREE OR NET OF LIFE

The emergence of modern eukaryotes from an endosymbiosis of organisms belonging to different phylogenetic groups demonstrates that the paradigm of a tree like

representation of the evolution is questionable. Horizontal gene transfer has also been discussed as an explanation for discrepancies between gene and species trees in the following cases: ATPase subunits of *Thermus thermophilus* (Gogarten *et al.*, 1992), *Methanococcus barkeri* and *Enterococcus hirae* (Hilario and Gogarten, 1993), glyceraldehyde phosphate dehydrogenase from *Escherichia coli* (Doolittle *et al.*, 1990), heat shock proteins (Gupta and Golding. 1993), glucose-6-phosphate isomerase (Smith and Doolittle, 1992), 16S rRNA of plant mitochondria (Gray *et al.*, 1989), P-elements in *Drosophila* (Daniels *et al.*, 1990) and glutamine synthases (Tiboni *et al.*, 1993). For a critical evaluation of some of these cases see Smith *et al.* (1992). It appears that horizontal transfer of genes between species is not restricted to resistance genes in modern-day bacteria, but occurred throughout evolution.

However, different molecular phylogenies that provide a robust resolution for the deep branches (as measured for example by bootstrap analyses) reveal identical or very similar topologies (e.g., 16S-like rRNAs, Woese, 1987; ATPases, Gogarten *et al.*, 1989; RNA polymerases, Puhler *et al.*, 1989; elongation factors, Cammarano *et al.*, 1992). This congruence of several molecular markers indicates that horizontal exchange of genetic information across species boundaries does occur only infrequently. Cases of horizontal transfer can be recognized within the background of the majority consensus of molecular markers. The fusion of separate lineages (net) is revealed by the simultaneous horizontal transfer of several independent genes (e.g.: the eucaryal cell organelles that evolved from bacterial endosymbionts).

Scenarios that Explain the Absence of Deep Branches

A common characteristic of the above mentioned molecular phylogenies (i.e., 16S-like rRNAs, ATPases, RNA polymerases, elongation factors), is that the two prokaryotic domains are clearly separated from each other. Although continuously deeper branching prokaryotes are discovered, so far, these clearly fall into one of the two domains.

One obvious possibility to explain the absence of deep branches in the rooted tree of life is chance. Different lineages terminate at random; it might be coincidence that so far none of the deep branches were detected, or they might all have gone extinct (see Figure 1a). As was pointed out by Zillig *et al.* (1992), this random death scenario is unlikely. More popular scenarios involve biological reasons as to why there were no or only a few deep branches in the first place (Figure 1b). Zillig *et al.* (1992) proposed a progenote population with rapid exchange of genes between members of the population. A first separation into two sub-populations is brought about by geographic isolation. Within each of the two sub-populations the rapid exchange of genetic information continues and prevents further speciation. The development of properties necessary for speciation (i.e. the step from a progenotic to a prokaryotic stage) occurred independently in both branches and has to be

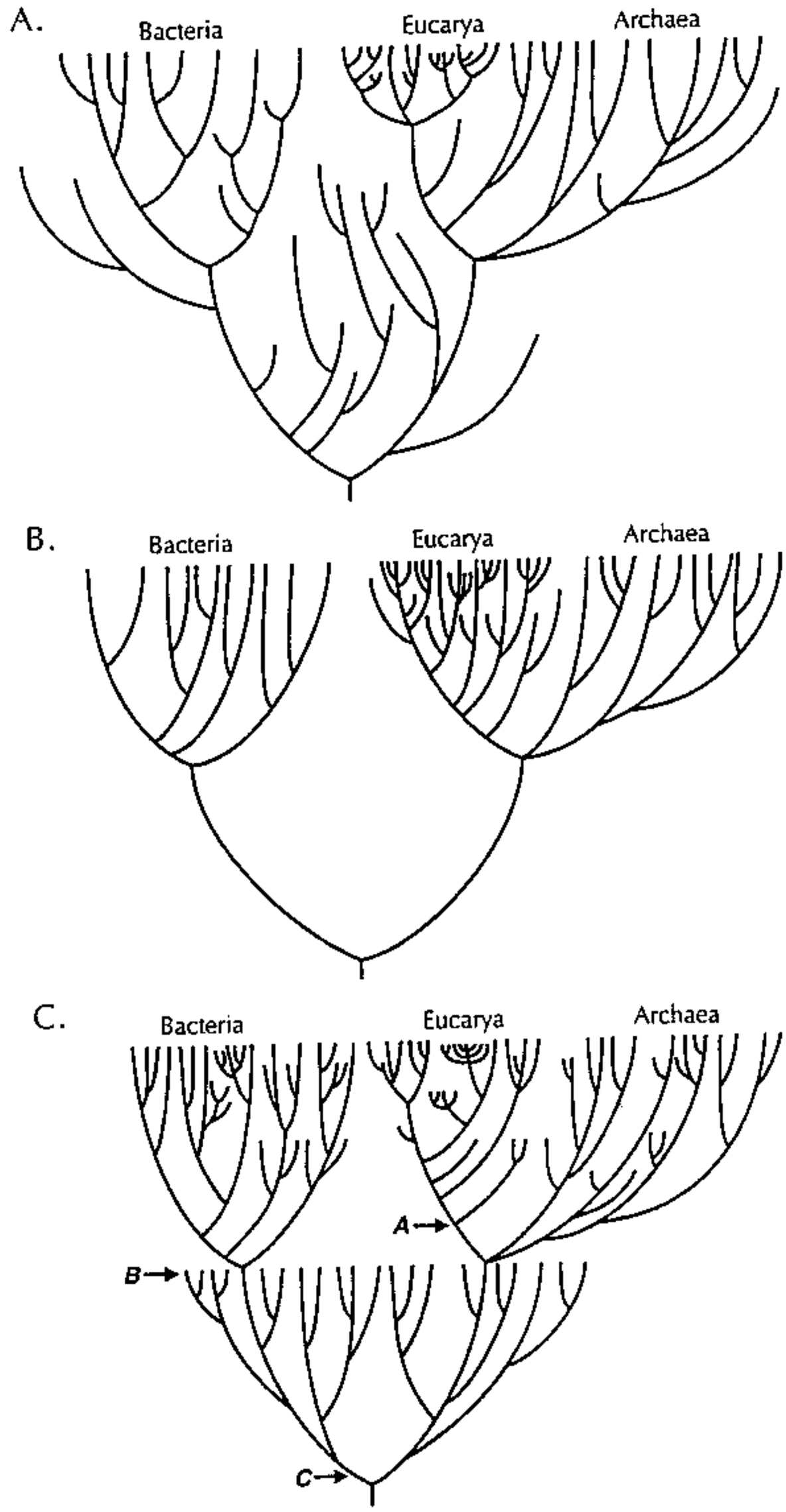

Fig. 1. The absence of deep branching surviving lineages in the molecular record can be explained by different scenarios. The phylogenetic tree depicted in panel A assumes a random distribution of extinguished lineages; panel B assumes that no deep branching lineages were generated in the first place; panel C depicts the proposed catastrophic extinction. The branch lengths represent an approximation of the number of substitutions that were calculated from the analyses of ancient duplicated genes (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989). The branches are not scaled with respect to time. Other molecular markers (e.g., 16S rRNA, Olsen and Woese, 1993; RNA polymerases, Puhler *et al.*, 1989) provide only unrooted molecular phylogenies; however, also in these cases the three domains are clearly separated from each other by long central branches. A, B, and C denote the origin of the Eucarya, the proposed catastrophic extinction and the last common ancestor, respectively. See text for further discussion.

regarded as a case of parallel evolution. One progenote population supposedly evolved into the Bacteria, the other into the Archaea.

A similar proposal was outlined by Kandler (1993). Again a population with rapid exchange of genetic information is assumed. Within this 'precell' population inventions are passed on between individuals. The ancestors of the three domains are thought to have separated from this pre-cell population sequentially. The differences between the three domains are explained by the following two processes: (1) Different subsets of genes from the precell population find their way into the three domain ancestors. (2) Some inventions are made in the precell population after one or two of the domain ancestors have already separated from the precell population.

Koch (1993) described a monophyletic phase during which no or only a few side branches were generated. Evolution supposedly occurred slowly and life did not adapt to new ecological niches. As the different life forms coexisted in the same ecological niche, all organisms were direct competitors which led to only a few survivors. Only after life adapted to different ecological niches did the different life forms avoid direct competition and separate surviving lineages formed.

Alternatively, the absence of independently surviving deep branches in the tree of life can be explained by a catastrophic event that wiped out all but two surviving lineages (Figure 1c). The two lineages surviving the catastrophe evolved into the Bacteria and the Archaea. The latter also contributed most of the eukaryotic nucleocytoplasm. In contrast to the other scenarios, the assumption of a catastrophic event eradicating most of the lineages from the bottom part of the tree does not assume a progenotic or primitive organizational level for the last common ancestor and the organisms that populated the bottom portion of the tree of life. Only the catastrophic extinction scenario is compatible with an already sophisticated prokaryotic last common ancestor.

The Last Common Ancestor

Molecular biology reveals the fundamental unity of modern life. All extant organisms are cellular, the genetic information is stored in DNA, transcribed into RNA, and translated into proteins. All organisms use the same (or very similar) genetic code and they use the same amino acids in their proteins. Although there are differences in the transcription and translation machinery, the process is very similar in all cells. All cells use lipid membranes to separate their protoplasm from the environment or from the cell wall; they use the same energy rich metabolites; and all living organisms use homologous enzymes to energize their cell membranes.

Furthermore, the study of ancient gene duplications shows that the last common ancestor already possessed a variety of complex enzymatic and regulatory processes. The membrane energizing ion translocation ATPases were already multi-subunit enzymes consisting of an ATP hydrolyzing subunit and a paralogous (i.e., derived by a gene duplication from the same ancestral gene) regulatory subunit. Structure

function relationships for extant ATPases (Cross and Taiz, 1990) suggest that the ATPases present in the last common ancestor already utilized transmembrane ion gradients for ATP synthesis (Gogarten and Taiz, 1992). Furthermore, the last common ancestor already had different paralogous elongation factors (Cammarano *et al.*, 1992), two types of ^{met}tRNA (Iwabe *et al.*, 1989), two glutamine synthases (Kumada *et al.*, 1993; Tiboni *et al.*, 1993), malate and lactate dehydrogenases (Iwabe *et al.*, 1989; Zillig *et al.*, 1992), an internal duplication in the carbamylphosphate synthetase (Puente, Gogarten, Lazcano, unpublished), different heat shock protein homologues (Gupta and Singh, 1992), and two glutamate dehydrogenases (Benachenhou-Lafha, 1993).

The picture of the last common ancestor emerging from the many shared characters of extant life and from the analysis of ancient duplicated genes is different from that of a primitive progenote. The last common ancestor appears to have been a prokaryotic cell that used DNA, RNA, ribosomes, energy conserving membranes and a variety of sophisticated, regulated biochemical pathways. The last common ancestor does not seem to have been fundamentally different from present day prokaryotes.

There is no *a priori* reason to assume that this cellular organism would have had difficulties undergoing speciation or adapting to different ecological niches present on the early Earth. The study of the recent evolution shows it is not a smooth continuous process, but characterized by major catastrophic events (Wilson, 1992). If catastrophic events shaped the evolution during the last 800 million years, we have every reason to assume that the same or similar forces also influenced the early evolution.

A Catastrophe Theory

The cratering record of the moon suggests that large impacts were likely to completely frustrate the development of life on Earth between 4000 and 3700 million years BP (Overbeck and Fogleman, 1989). The last impact with an energy sufficient for sterilizing the Earth (i.e., to completely vaporize the oceans) is estimated to have occurred between 4400 and 3800 million years BP; impacts vaporizing the photic zone only were calculated to be highly probably as late as 3800 million years BP (Sleep *et al.*, 1989). A large impact would have drastically altered or obliterated most of the available ecological niches. Such an impact appears as a likely cause for the catastrophe that is suggested by our interpretation of the molecular records. In this context it is of significance that the different proposals for the universal tree of life (Lake, 1988; Woese, 1987) agree in the deepest branches of the respective trees being occupied by extreme thermophiles. Only extremely thermophilic organisms would have survived a nearly ocean boiling impact. As was also pointed out by Lazcano (1993b), the distribution of thermophiles on the tree of life is not necessarily indicative of life having originated at high temperature (Pace, 1991; Holm, 1992), it might be reflective of the selection that took place during the peri-

od of heavy bombardment. The idea of life's origin at high temperature has also been criticized because of the instability of biomolecules at elevated temperatures (Miller and Bada, 1988), and because an enzyme that seems to be necessary for DNA stability at high temperatures (reverse gyrase) appears to have evolved only later during evolution (Forterre *et al.*, 1993).

The described catastrophic impact scenario reconciles the above cited criticisms with the distribution of extreme thermophiles on the tree of life. Life could have originated in a mesophilic environment. Prokaryotic life already had diversified and occupied the different ecological niches available on the planet before the occurrence of the late catastrophic impact. Among other niches prokaryotes also had settled the extremely hot environments of the deep ocean vents. Only those prokaryotes that had adapted to hot environments were able to survive a large impact.

Dating the Molecular Record

There is no justification to assume a molecular clock that is running at constant speed throughout evolution. To the contrary, the study of the more recent molecular evolution, in particular of vertebrates, showed up to 13-fold transient increases in substitution rates after a gene duplication event had occurred (Ohta, 1991 and ref. therein). The root in the universal tree of life was placed by means of ancient gene duplications, therefore, it seems justified to assume that the bottom part of the trees outlined in Figure 1 is enlarged with respect to the upper portions. (Note, however, that the deep branches calculated from 16S rRNA are only slightly shorter than the ones depicted in Figure 1; cf. Olsen and Woese, 1993). On the other hand, the character of the last common ancestor that is inferred from ancient duplicated genes offers no indication for the substitution rate in the bottom portion of the tree being higher by several orders of magnitude. Only few events are available for an attempt to calibrate the molecular record by means of the fossil record.

THE EMERGENCE OF EUKARYOTES

Eukaryotic cells are present in the fossil record since about 1750 million years BP. Sediments from about 2000 million years BP contain fossils of likely eukaryotic origin (Schopf, 1992). The characterization of microfossils as eukaryotic is based mainly on cell size; therefore, it appears likely that the earliest eukaryotes, even if they are present in the fossil record will not be classified as such. Our conservative estimate for the separation of the eukaryotic nucleocytoplasmic lineage from the archaebacteria is about 2000 million years BP. However, it is likely that the independent eukaryotic lineage is significantly older than this estimate. The molecular record indicates that the separation between Archaea and Eucarya must have occurred after the postulated impact. Therefore, we can conclude that the late, nearly complete impact frustration of life must have occurred around or well before 2000 million years BP.

EARLY ARCHEAN MICROFOSSILS

Microfossils have been described in rocks dating back to around 3500 million years BP. The structurally best conserved samples are from the Pilbara rocks (Schopf, 1993) and the Warrawoona group, both in Western Australia (cf. Awramik, 1992). Microfossils were also described in rocks from the Swaziland supergroup in South Africa (Walsh, 1992). Stromatolites were described in the Swaziland supergroup (Byerly *et al.*, 1986) and the Warrawoona group (cf. Awramik, 1992). Modern stromatolites are the result of interactions of complex microbial communities, including oxygen producing photosynthetic cyanobacteria. The morphologies of some of the early microfossils are similar to extant *Oscillatoria*. This morphological resemblance and the presence of stromatolites were interpreted as proof (Awramik, 1992) or at least strongly suggestive of cyanobacteria being present already about 3500 million years BP. While some of these findings have been questioned (summarized by Schopf, 1993), the resemblance between filamentous microfossils and present day cyanobacteria is impressive. However, other bacterial groups (e.g., the green non sulfur bacteria *Chloroflexus, Heliothrix* and *Oscillochloris* and the non-photosynthetic Beggiatoales) contain similar multicellular filamentous forms as well (Pfennig, 1989; Strohl, 1989); some of these also resemble *Oscillatoria*. If we assume that the early Archaean cyanobacteria-like microfossils are representatives of lineages that went extinct during the catastrophic impact, then these microfossils cannot be used for calibration. However, if we consider them to be representatives of surviving bacterial groups, then an accelerated substitution rate for the bottom portion of the tree has to be assumed in order to accommodate the tree of life within the life time of this planet.

The properties of the last common ancestor and the study of more recently duplicated genes (see the above discussion) suggest that the substitution rate in the bottom portion of the tree was temporarily increased by less than thirteen fold. Assuming a tenfold higher substitution rate throughout the bottom portion of the tree, the last common ancestor (C in Figure 1C) is to be dated to about 3900 million years BP; the point of the catastrophic extinction (B in Figure 1C) is estimated to have occurred between 3600 and 3700 million years BP. If a less accelerated substitution rate is assumed, even earlier times result for the last common ancestor and the nearly complete impact frustration.

The Prokaryote – Eukaryote Dichotomy

As discussed above many molecular markers reflect the deep separation between Archaea and Bacteria. Considering cell biological characters the distinction between the Pro- (Archaea and Bacteria) and the Eukaryotes (Eucarya) appears at least as fundamental (compare the discussion in Mayr, 1990). This basic distinction is also reflected in the first molecular phylogenies. The branch that connected the eucaryal domain to the two prokaryotic domains was at least as long as the branch that led

from the central trifurcation (i.e., the node in which the three domains converge in an unrooted phylogeny) to the Bacteria (e.g.: Olsen, 1987).

However, Sogin *et al.* (1989) reported that the 16S like rRNA from the protist *Giardia lamblia* contains many signature residues typical for prokaryotes. *Giardia lamblia* is a parasitic flagellated protist with two nuclei. *Giardia* has a reduced endomembrane system that shares many functional features with other Eukaryotes (e.g.: a regulated secretory pathway; Reiner *et al.*, 1990). The finding that *Giardia* represents one of the deepest branching eucaryal lineages (Sogin *et al.*, 1989) suggests that many of the primitive features (e.g.: no mitochondria, reduced ER and Golgi) might be primary features and not adaptations to a parasitic life style.

The following finding complicated the analysis of signature of the 16S rRNA: the microsporidian *Vairimorpha necatrix* branches off nearly as deep or even deeper than *Giardia*; however, the *Vairimorpha* sequence does not contain many prokaryotic signatures (Table 1 in Sogin *et al.*, 1989). One peculiarity of the 16S rRNA from *Giardia* is its high GC contents (75%). In contrast the GC contents of protein encoding genes in *Giardia* ranges between 49 and 60% GC (Adam, 1991).

Using PCR we obtained a fragment of the vacuolar ATPase *A*-subunit, which in turn was used for screening a *Giardia lamblia* genomic library in λ_{ZapII} (kindly provided by Dr. Frances Gillin, UC San Diego). The encoded protein sequence is intermediate between the prokaryotic and the eukaryotic sequences (Figure 2). The pro- and eukaryotic consensus sequences for the depicted region differ in two positions (marked by #). The *Giardia* sequence contains the prokaryotic signatures in these positions. Five additional positions of the *Giardia* sequence have amino acid residues that are found in only one of the two consensus sequences. These positions are indicated by arrows in the Table. In a total of five positions does the *Giardia* sequence agree exclusively with the prokaryotic consensus, in only two positions does the *Giardia* sequence reflect eukaryotic specific signatures. Although *Giardia* is a true Eukaryote, concerning the primary structure of its macro-molecules *Giardia* appears to be halfway between the other Eukaryotes and the Prokaryotes. Concerning some molecular markers the separation between Pro- and Eukaryotes is less pronounced than the one between Archaea and Bacteria. So far this statement appears to be true for the 16S-like rRNA and the catalytic subunit of the vacuolar type ATPase. In case of the 16S rRNA the analysis is complicated by the high GC content of the *Giardia* sequence and by the apparently accelerated substition rate within the eucaryal lineage.

The fundamental difference between pro- and eukaryotes lies in their cell-biological differences. However, these cell biological achievements were obtained over a short evolutionary distance (as measured in number of substitution events in biological macromolecules).

```
Daucus          FRDMGYNVSMMADSTSRWAEALREISGRLAEMPADSGYPAYLAARLASFYERAG
Human           FRDMGYNVSMMADSTSRWAEALREISGRLAEMPADSGYPAYLGARLASFYERAG
Sus             FRDMGYHVSMMANSTSRWAEALREISGRLAEMPADSGYPAYLGARLASFYERAG
Manduca         FRDMGYNVSMMADSTSRWAEALREISGRLAEMPADSGYPAYLGARLASFYERAG
S. pombe        YRDQGKNVSMMADSTSRWAEALREISGRLAEMPADSGYPAYLGAKLASFYERAG
Candida         FRDQGKNVSMIADSSSRWAEALREISGRLGEMPADQGFPAYLGAKLASFYERAG
Neurospora      FRDQGMNVAMMADSSSRWAEALREISGRLGEMPADQGFPAYLGAKLASFYERAG
Trypanosom.     YRDMGKHIAMMADSTSRWAEALREISGRLAEMPADGGYPAYLSARLASFYERAG
Plasmodium      FRDMGYNATMMADSTSRWAEALREISGRLAEMPADSGYPAYLGARLASFYERAG

eukaryotic      Y M N     I D S                G       Y       R
                FRDQG H   MMANSTSRWAEALREISGRLAEMPAD GFPAYL AK LASFYERAG
                ***** *   ********************  *  **** *  **  ******

Giardia         FRDQGYNVTLLADSTSRWAEALREISGRLEGIPGEGGYPADLTSKLSHFYERAG
                *****  * * ************ *****  *** ***  * * ** ******

prokaryotic     FYRDMQG  V LVIMADSTSRWAEALMRDE SGRLMEEMPGEADEGFYP YL ASRLSA EYERASG

Halobact.       YRDMGYDVALMADSTSRWAEAMREISSRLEEMPGEEGYPAYLAARLSEFYERAG
Methanos.       YRDMGLDVSLMADSTSRWAEAMREISSRLEEMPGEEGYPAYLSARLAEFYERAG
Methanococc.    FRDMGYGVLLTADSTSRWAEAMREISGRLEEMPGEEGYPAYLASRLAQFYERAG
Sulfolobus      FRDQGYDVLLVADSTSRWAEALRDLGGRMEEMPAEEGFPSYLPSRLAEYYERAG
Thermus         FRDQGFSVALMADSTSRWAEALREISSRLEEMPAEEGYPPYLAARLAAFYERAG
E. hirae        FRDMGYDVAIMADSTSRWAEALREMSGRLEEMPGDEGYPAYLGSRLAEYYERSG
                          #                   #
                ↑   ↓                       ↓   ↓↓          ↓↑ ↓
```

Fig. 2. Alignment of partial amino acid sequences of the catalytic subunit of vacuolar type ATPases from various pro- and eukaryotes. 'Eukaryotic' and 'prokaryotic' denote the respective consensus sequences. The *Giardia lamblia* sequence was not considered for the eukaryotic consensus. For a given position a consensus residue is only given, if less than two amino acids are used in this position. Matches between the *Giardia* sequence and the two consensus sequences are indicated by '*'. '#' denotes positions in which the prokaryotic and the eukaryotic consensus sequences differ. Positions in which the *Giardia* sequence corresponds to only the eukaryotic (↑) or only the prokaryotic (↓) consensus sequence are indicated by arrows. Accession numbers for the sequences in the GenBank peptide (gp) and PIR databank are as follows: *Daucus carota* ,gp|J03769; Human ,gp|L09234; *Sus scrofa* ,gp|X62338; *Manduca sexta* ,gp|X64233; *Schizosaccharomyces pombe* ,gp|X68580; *Candida tropicalis* ,gp|M64984; *Trypanosoma congolense* ,gp|Z25814; *Plasmodium falciparum* ,gp|L08200; *Halobacterium salinarium* ,gp|X70294; *Methanosarcina barkeri* ,gp|J04836; *Methanococcus thermolithotropicus* ,PIR|S13589; *Sulfolobus acidocaldarius* ,gp|J03218; *Thermus thermophilus* ,gp|X63855; *Enterococcus hirae* ,gp|D13816. For the depicted portion the corresponding sequence fragment from *Saccharomyces cerevisiae* (gp|M21609) is identical to the *Candida* sequence; the second Human (gp|L09235) and the Bovine (gp|M80430/X58386) sequences are identical to the depicted fragment of the Human sequence; and the sequence from cotton (*Gossypium hirsutum*; gp|L03186) is identical to the depicted sequence from carrot (*Daucus carota*).

Predictions and Implications Derived from the Impact Scenario

The assumption of a catastrophic, nearly complete extinction of an already diversified complex prokaryotic biota makes it unlikely that surviving missing links between the Archaea and the Bacteria will be found. However, the discovery of independent deep branching lineages can not be ruled out; after all, two lineages have survived the period of the early heavy bombardment.

It is already difficult to prove the involvement of impacts in the much more recent mass extinctions at the Cretaceous-Tertiary and the Permian-Triassic boundaries. The Precambrian, in particular the Archaean, fossil and sedimentary records are significantly sparser than those from the Paleo- and Mesozoic eras. Accordingly, the task of accumulating geological and fossil evidence pointing to specific catastrophic events during the Archaean is even more difficult. However, because of the expected greater magnitude of the Archaean catastrophes an even slightly more detailed microfossil record than the one presently available can be expected to result in sufficient continuity to more closely pinpoint early dramatic evolutionary events.

The proposed catastrophic extinction scenario is compatible with an extremely thermophilic last common ancestor and with an origin of life at high temperatures. However, the proposed mass extinction by meteorite impact also provides a strong selection pressure that could have selected for extreme thermophily. A corollary to the interpretation of extreme thermophily as a derived character is that adaptation to extremely high temperatures is likely to have occurred independently in the two lineages leading to the Bacteria and Archaea. The hypothesis of extreme thermophily as a derived character would be verified if more detailed molecular studies would show an independent and parallel evolution of extreme thermophily in Archaea and Bacteria. The reverse is not necessarily true: homologous enzymes that are responsible for the adaptation to extremely high temperatures in Archaea and Bacteria might also be due to horizontal gene transfer, and not due to an extreme thermophily of the last common ancestor. To falsify extreme thermophily as a derived character a more detailed phylogenetic analysis of the pertinent enzymes is necessary; in particular, the congruence of the molecular phylogenies with other molecular markers is a prerequisite.

Another approach to verify/falsify the nature of extreme thermophily as a derived character is to study the evolution of enzymes, metabolic pathways and physiological responses that play key roles in adaptation to high temperature. If it is shown that these traits were assembled from precursors that are essential to cellular function but in themselves do not provide or contribute to an adaptation to high temperatures, then it has to be concluded that these basic cellular functions evolved in mesophilic and not in thermophilic organisms. The case of reverse gyrase, which is suggested to have evolved from a fusion of a topoisomerase and a helicase (Confalonieri *et al.*, 1993), can be regarded as a first example (Forterre *et al.*, 1993).

Conclusion

The described catastrophic extinction scenario is compatible with the molecular, fossil and astrophysical records. It suggests that life emerged early in Earth's history even before the time of the heavy bombardment was over. Early life forms already had colonized extreme habitats that allowed at least two prokaryotic species to survive a late nearly ocean boiling impact. The distribution of ecotypes on the

rooted universal tree of life cannot necessarily be interpreted to show that life originated in extremely hot environments.

Acknowledgments

We thank Henrik Kibak, Lincoln Taiz, Antonio Lazcano and Wolfram Zillig for many stimulating discussions, Frances Gillin for providing us with a genomic library from *Giardia lamblia*, and Scot Magnotta for critically reading the manuscript. This work has been supported through NSF grant BSR-9020868. EH was supported in part by a fellowship from the DGAPA-Universidad Nacional Autónoma de México.

References

Adam, R. D.: 1991, *Microbiol. Rev.* **55**, 706–732.
Awramik, S. M.: 1992, *Photosynth. Res.* **33**, 75–89.
Benachenhou-Lafha, N., Forterre, P. and Labedan, B.: 1993, *J. Mol. Evol.* **36**, 335–346.
Bruns, T. D., White, T. J. and Taylor, J. W.: 1991, *Annu. Rev. Ecol. Syst.* **22**, 525–564.
Byerly, G. R., Lowe, D. R. and Walsh, M. M.: 1986, *Nature.* **319**, 489–491.
Confalonieri, F., Elie, C., Nadal, M., Bouthier de la Tour, C., Forterre, P. and Duguet, M.: 1993, *Proc. Natl. Acad. Sci.* **90**, 4753–4757.
Cammarano, P., Palm, P., Creti, R., Ceccarelli, E., Sanangelantoni, A. M. and Tiboni, O.: 1992, *J. Mol. Evol.* **34**, 396–405.
Cross, R. L. and Taiz, L.: 1990, *Fed. Eur. Biochem. Soc. Lett.* **259**, 227–229.
Daniels, S. P., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G. and Chovnick, A.: 1990. *Genetics* **124**, 339–355.
Doolittle, R. F., Feng, D. F., Anderson, K. L. and Alberro, M. R.: 1990, *J. Mol. Evol.* **31**, 383–388.
Eigen, M., Lindemenn, B. F., Tietze, M., Winkler-Oswatitsch, R., Dress, A. and von Haeseler, A.: 1989, *Science* **244**, 673–679.
Forterre, P., Confalonieri, F., Charbonnier, F., Bouthier de la Tour, C. and Duguet, M.: 1993, Proceedings of the 10th International Conference on the Origin of Life, 184.
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, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6661–6665.
Gogarten, J. P. and Taiz, L.: 1992, *Photosynth. Res.* **33**, 137–146.
Gogarten, J. P. Starke, T., Kibak, H., Fichmann, J. and Taiz, L.: 1992 *J. Exp. Biol.* **172**, 137–147.
Granick, S.: 1957, *Ann. NY Acad. Sci.* **69**, 292–308.
Gray, M. W., Cedergren, R., Abel, Y. and Sankoff, D.: 1989, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2267–2271.
Gupta, R. S. and Singh, B.: 1992, *J. Bacteriol.* **174**, 4594–4605.
Gupta, R. S. and Golding, G. B.: 1993: *J. Mol. Evol.* **37**, 573–582.
Hilario, E. and Gogarten, J. P.: 1993, *Biosystems* **31**, 111–119.
Holm, N. G. (Ed.): 1992, *Marine Hydrothermal Systems and the Origin of Life*. Kluwer, Norwell, MA. U.S.A.
Iwabe, N., Kuma, K-I., Hasegawa, M., Osawa, S. and Miyata, T.: 1989, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9355–9359.
Kandler, O.: 1993, *Early Life on Earth*: Nobel Symposium No. 84. S. Bengston (Ed.), Columbia University Press, New York, pp. 111–118.
Koch, A. L.: 1993, Proceedings of the 10th International Conference on the Origin of Life, 140.
Kumada, Y., Benson, D. R., Hillemann, D., Hosted, T. J., Rochford, D. A., Thompson, C. J., Wohlleben, W. and Tateno, Y.: 1993, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3009–3013.
Lake, J. A.: 1988, *Nature* **331**, 184–186.

Lazcano, A.: 1993a, *Early Life on Earth*: Nobel Symposium No. 84. S. Bengston (Ed.), Columbia University Press, New York, pp. 59–80.
Lazcano, A.: 1993b, *Science* **260**, 1154–1155.
Mahler, K. A. and Stevenson, D. J.: 1988, *Nature* **331**, 612–614.
Margulis, L.: 1981, *Symbiosis in Cell Evolution*, W. H. Freeman and Company, New York.
Mayr, E.: 1990, *Nature* **348**, 491.
Moran, N. A., Munson, M. A.., Baumann, P. and Ishikawa, H.: 1993, *Proc. R. Soc. Lond. B* **253**, 167–171.
Miller, S. L. and Bada, J. L.: 1988, *Nature* **334**, 609.
Miller, S. L.: 1992, *Major Events in the History of Life*, J. W. Schopf. (Ed.), Jones and Bartlett Publ., Boston, pp. 1–28.
Ochman, H. and Wilson, A. C.: 1987, *J. Mol. Evol.* **26**, 74–86.
Ohta, T.: 1991, *J. Mol. Evol.* **33**, 34–41.
Olsen, G., Cold Spring Harb. Symp. Quant. Biol.: 1987, **52**, 825–837.
Olsen, G. L. and Woese, C. R.: 1993, *FASEB Journal* **7**, 113–123.
Overbeck, V. R. and Fogleman, G.: 1989, *Nature* **339**, 434.
Overbeck, V. R. and Fogleman, G.: 1990, *Origins of Life and Evolution of the Biosphere* **20**, 181–195.
Pace, N. R.: 1991, *Cell* **65**, 531–533.
Pääbo, S.: 1993, *Sci. Am.* **263**, 5, 86–92.
Pfennig, N.: 1989, in *Bergey's Manual of Systematic Bacteriology*, Vol. **3**, J. T. Staley, M. P. Bryant, N. Pfennig and J. G. Holt. (Eds.), Williams and Wilkins, Baltimore, pp. 1697–1707.
Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A. and Zillig, W.: 1989, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4569–4573.
Reiner, D. S., McCaffery, M. and Gillin, F. D.: 1990, *Eur. J. Cell Biol.* **53**, 142–153.
Schopf, J. W.: 1992, *Major Events in the History of Life*, J. W. Schopf. (Ed.), Jones and Bartlett Publ., Boston, pp. 54–63.
Schopf, J. W.: 1993, *Science* **260**, 640–646.
Schwartz, R. M. and Dayhoff, M. O.: 1978, *Science* **199**, 395–403.
Sleep, N. H., Zahnle, K. J., Kasting, J. F. and Morowitz, H. J.: 1989, *Nature* **342**, 139–142.
Smith, M. W. and Doolittle, R. F.: 1992, *J. Mol. Evol.* **34**, 544–545.
Smith, M. W., Feng, D. F. and Doolittle, R. F.: 1992, *Trends Biochem. Sci.* **17**, 489–493.
Sogin, M., Gunderson, J. H., Elwood, H. J., Alonso, R. A., Peatie, D. A.: 1989, *Science* **243**, 75–77.
Sogin, M. L.: 1992, *Curr. Opinion Genet. Devel.* **1**, 457–463.
Strohl, W.: 1989, in *Bergey's Manual of Systematic Bacteriology*, Vol. **3**, J. T. Staley, M. P. Bryant, N. Pfennig, and J. G. Holt. (Eds.), Williams and Wilkins, Baltimore, pp. 2089–2105.
Tiboni, O., Cammarano, P. and Sanangelantoni, A. M.: 1993, *J. Bacteriol.* **175**, 2961–2969.
Walsh, M. M.: 1992, *Precambrian Res.* **54**, 271–293.
Wächtershäuser, G.: 1990, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 200–204.
Wilson, E. O.: 1992, *The Diversity of Life*, Harvard University Press, Cambridge, MA, U.S.A.
Woese, C. R. and Foc, G. E.: 1977a *J. Mol. Evol.* **10**, 1–6.
Woese, C. R. and Fox, G. E.: 1977b, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088–5090.
Woese, C. R.: 1987, *Microbiol. Rev.* **51**, 221–271.
Woese, C. R., Kandler, O. and Wheelis, M. L.: 1990, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579.
Zillig, W., Palm, P. and Klenk, H-P.: 1992, *The Origin and Evolution of the Cell*, H. Hartman, K. Matsuno. (Eds.), World Scientific, Singapore, New Jersey, London, Hong Kong, pp. 163–183.