

Genetics Indicates Extra-terrestrial Origins for Life: The First Gene

Did Life Begin Following the Big Bang?

Rhawn Joseph¹ and N. Chandra Wickramasinghe²

¹Emeritus, Brain Research Laboratory, Northern California.

² Buckingham Centre for Astrobiology (BCAB), The University of Buckingham,
Buckingham MK181EG UK

Abstract

Some of the defining characteristics of life include its ability to replicate and reproduce itself and its genome whilst maintaining a capacity for evolution. Life, as we know it, requires genetic information and no fewer than 382 genes. Thus, the origins of life can be estimated by determining duplication rates beginning with the first gene. Conversely, an approximate date for the origin of the first gene can be estimated based on the time frame in which life appeared on Earth and then evolved. Data from genomics and molecular biology indicate that all modern organisms originated from an ensemble of prokaryotic genes dating back to the first appearance of life on Earth over 4.2 billion years ago (bya), which means life was present on Earth from almost the beginning. There are two main models for the origins of eukaryotes on Earth, “genetic merger” which postulates that eukaryotes evolved between 3 bya to 2 bya following the merger of two species of prokaryote, and “deep roots” which posits that eukaryotes and prokaryotes appeared on Earth at the same time. However, neither model can explain how or when the first genes evolved. To arrive at an estimate for the time frame in which the first gene was fashioned, genetic analysis based on the genetic merger and deep roots models and at different ages for the establishment of life on Earth, were carried out. The correspondence between total gene numbers of various organisms and time of their putative origin over the course of evolutionary history was analysed. Be it the merger or deep roots model, and within the experimental uncertainties in the relevant data, four separate analyses indicates that the origins of the first gene extends backwards in time by an estimated 10 to 14 billion years ($10.5 < 14.5$ bya) and thus during a time frame which overlaps and is consistent with estimated ages offered in support of the Big Bang model of the origin of this universe. This does not mean that life began 10-14 billion years ago, but rather that the first gene was fashioned billions of years before the creation of Earth.

Keywords: Genetics, Origins of Life, Abiogenesis, Genome, Duplication, Replication, Viruses, RNA Worlds

Introduction

When life first appeared on Earth is unknown. Likewise, the birth date for the creation of the first gene is as yet undetermined. Although definitive fossil evidence is lacking, there are tantalizing clues which suggest life may have been present on Earth, fractionating and synthesizing carbon as early as 3.8 to 4.28 billion years ago (Manning et al. 2006; Mojzsis et al. 1996; Nemchin et al. 2008; O'Neil et al. 2008). Specifically, microprobe analyses of the carbon isotope composition of metasediments in Western Australia formed 4.2 billion years ago (bya) has revealed very high concentrations of carbon 12, or "light carbon" which is typically associated with microbial life (Nemchin et al. 2008). The discovery of banded iron formations in northern Quebec, Canada, consisting of alternating magnetite and quartz dated to 4.28 bya, may also be associated with biological activity (O'Neil et al. 2008). In addition, carbon-isotope analysis of a phosphate mineral, apatite, in quartz-pyroxene rocks on Akilia, West Greenland and dated to 3.8 bya, was found to contain tiny grains of calcium and high levels of organic carbon; which is suggestive of photosynthesis, oxygen secretion, and thus biological activity (Manning et al. 2006; Mojzsis et al. 1996). The high carbon contents of the protolith shale from S. W. Greenland, and the ratio of carbon isotopes in graphite from metamorphosed sediments dating to 3.8 bya is also suggestive of photosynthesizing activity (Rosing, 1999, Rosing and Frei, 2004).

Genetic data based on molecular clocks is consistent with these discoveries. According to analysis performed by Feng et al., (1997) and Hedges (2001) single celled prokaryotes were sharing the planet 4 billion years ago. Based on an analysis of the eukaryotic genome, Hedges (2001) reports "we found an early time of divergence (approximately 4 billion years ago, Ga) for archaeobacteria and the archaeobacterial genes in eukaryotes." Hedges (2009) concludes, "life on Earth arose... 4400 to 4200 million years ago, and achieved a prokaryote level of complexity. An initial split, 4200 led to Superkingdoms Eubacteria and Archeobacteria."

There is thus a confluence of evidence suggesting that life was present on this planet during a period known as the "Late Heavy Bombardment" (Schoenberg et al. 2002) when Earth was pummeled with massive extra-terrestrial debris causing the surface to melt and form new rocky layers. Under these conditions, all evidence of life prior to 4.2 bya ago would have been obliterated. Likewise, any fossilized remains would have continued to be pulverized until after the end of the bombardment period, which may explain why the first definitive evidence for microbial life does not appear until 3.5 bya (Furnes et al. 2004). Since Earth is believed to have been formed 4.6 billion years ago, the genomic and biological evidence indicates life was present on Earth nearly from the very beginning of

this planet's formation. Presumably, these life forms were single celled organisms capable of biological activity as demonstrated by the biological fingerprints left in this planet's oldest rocks dated to 4.2 bya (Nemchin et al. 2008; O'Neil et al. 2008).

When the first gene was formed is unknown. Likewise, when eukaryotes first took up residence on Earth is a matter of great speculation and debate. However, the origins of eukaryotes can help unravel the mystery for the origins of the first gene.

There are two major views on the origins of Earthly eukaryotes. The genetic “merger models” as embraced by Hedges (2004, 2009) and others, posits that different species of bacteria (e.g. archeobacterium and a eubacterium) joined together thereby creating compartmentalization within the cell, and thus a nucleus around 2.7 bya, and then later, a mitochondrion around 2 bya to 2.4 bya (Joseph 2010a). By contrast, the “deep roots” models hold that eukaryotes appeared at the same time prokaryotes began to diverge (Hartman, & Fedorov, 2002; Kurland, Collins, & Penny, 2007; Poole, & Penny, 2007). If the deep roots model is correct, this would mean that eukaryotes and prokaryotes were present, on Earth, between 4.4 bya to 3.8 bya. In support of this later possibility is the controversial evidence suggestive of microfossils resembling yeast cells and fungi discovered in 3.8 billion year old quartz (Pflug 1978).

Woese (2004) has proposed that life began with proto cells and the initial proto cells may have lived together and repeatedly swapped and shared genes via horizontal gene transfer. "Eventually this collection of eclectic and changeable cells coalesced into the three basic domains known today" (Woese, 2004). According to this proposal, bacteria, archaea and eukaryotes may have been established at around the same time and differentiated from a diverse collection of proto-cellular ancestors. However, if proto-cells possessed genes, and were essentially life-like and capable of biological activity 4.2 to 4.4 bya, then where did they obtain their genes, and how did their genes acquire the ability to replicate?

Be it "deep roots" or "merger models", neither can supply an answer as to the origins of the first genes or the origins of life. Certainly, life, as we know it, is not possible without a genome and a minimal number of diverse, life sustaining genes. For example, the prokaryote, *Mycoplasma genitalium*, has the fewest genes of all living organisms, 485 genes vs 552 genes for *Nanoarchaeum equitans* (which has fewer base pairs, i.e. 490,885). *Mycoplasma* has also been shown to remain viable even after 100 of its genes were removed (Fraser et al., 1995; Glass et al., 2006). Thus, it requires at least 382 genes for a single cell to live and this suggests that the first life forms must have also first acquired a genome of this size. However, it takes at least 485 genes to survive and replicate. It is this

range in minimal gene numbers (382 to 485 genes) which provides us with a key to unlocking the origins of the genetic basis of life and the birth date of the first gene.

Certainly it is not plausible that 382 to 485 genes and a minimum of 490,885 base pairs were established *ex-nihilo* and simultaneously on Earth or any other environment and from which life immediately sprang. Therefore it can be assumed the first genes must have begun to evolve prior to the establishment of life, and it was only over the course of evolutionary development and subsequent gene and whole genome duplication that a minimal gene set necessary for life was established.

As discussed below, genes and the genome have been repeatedly duplicated over the course of evolution. Therefore, a determination of the rate of gene duplication which led to the establishment of life on Earth, can lead us backwards in time to an approximate date when the first genes were generated. As detailed in this article, and as based on a series of statistical analyses of the "deep roots" or "merger models", and of various dates for the origin of Earthly life (4.4 to 3.5 bya), we estimate that life has a genetic ancestry of approximately 10 billion years.

2. Genome Duplication

Some of the defining characteristics of life include its ability to replicate and reproduce itself and its genome whilst maintaining a capacity for evolution. Life, as we know it, requires genetic information and at a minimum nearly five hundred diverse genes in order to replicate and at least 382 genes just to survive.

It is widely accepted that the eukaryotic genome has increased in size over the course of evolution; a function of gene and whole genome duplication (WGD) as well as horizontal gene transfer (HGT). Likewise, it can be predicted that after the first genes were established, they must have acquired the ability to replicate, thereby increasing in number and becoming more variable until finally a core set of at least 382 genes were established, as this is the minimal number necessary for maintaining single cellular life (Fraser et al., 1995; Glass et al., 2006). These and other core genes, beginning with the first gene, were duplicated and continued to increase in number after prokaryotes and eukaryotes evolved. Duplication rates, therefore, can provide us with an approximate date for the creation of the first gene.

There have been repeated episodes of WGD during the early evolution of eukaryotes and which date back to the emergence of the first eukaryotic cells or their ancestors (Makarova et al., 2005). The eukaryote genome appears to have been duplicated every 100 million

years (Lynch et al., 2001; Lynch and Conery 2000), though in fact the frequency is as yet unknown. Whole genome duplications have occurred in almost all lineages, including yeast (Wong et al., 2002; Vision et al., 2000; Kellis et al., 2004), fish (Van de Peer et al., 2003; Jaillon et al., 2004; Taylor et al., 2001), frogs (Tymowska et al., 1977; Jeffreys et al., 1980) and plants (Blanc and Wolfe 2004). The relatively large and complex vertebrate genome appears to have been duplicated at least twice (McLysaght et al., 2002; Dehal and Boore 2005), suggesting a duplication rate of once every 250 million years.

Single gene and whole genome duplication played a central role in the primary radiation of chordates (Dehal and Boore 2005) during the Cambrian explosion of metazoan life 540 million years ago. There followed two additional duplications during chordate evolution, thereby forming many of the gene families of vertebrates (McLysaght et al., 2002). Thus it appears that two distinct genome duplication events occurred early in vertebrate evolution and after vertebrates began to colonize the surface of Earth (Dehal and Boore 2005), which again suggests a duplication rate every 250 million years.

However, duplication is often followed by accelerated sequence evolution as well as rearrangement of a gene, and even gene deletion; evolutionary modes that obliterate detectable connections to the original prokaryotic, proto-cell or viral gene source.

3. Genomic Duplication, Evolution, Divergence and the First Gene

The dozens of duplicative events over the course of evolutionary history (Lynch and Conery 2000; Lynch et al., 2001) has likely triggered the transition and divergence between numerous species, ranging from yeast and fungi (Liti and Louis, 2005) to chordates and non-chordates (Dehal and Boore 2005; McLysaght et al., 2002). Likewise, it can be assumed that duplicative events led to the divergence of archae and bacteria which is assumed to have taken place, on Earth, over 4 billion years ago, and previous duplicative events led to the divergence of prokaryotes from proto-cells.

It is also reasonable to assume that gene numbers and the genome must have doubled in size after the establishment of the first gene. These duplicative events continued until a minimal gene set necessary for life was established.

4. Genomic Data Indicates Genes Were Formed Prior to Creation of Earth

Genetic clocks indicate that life was present on Earth by 4.4 to 4.2 bya and included at least two super kingdoms of prokaryotes (Hedges 2001, 2009). These findings are supported by biological residue associated with life (Nemchin et al. 2008; O'Neil et al.

2008). If the first gene was established at the same time Earth was created, 4.6 bya, this would mean that the first gene (and the first genome) had to acquire the ability to replicate itself and to undergo at least 9 to 10 duplicative events, coupled with deletions and mutations, within a span of 200 million years to achieve a minimal gene set of 382 genes, by 4.4 bya. However, duplicative events in prokaryotes appear to occur over a span of billions of years, whereas whole genome duplicative events occur in the eukaryote genome perhaps once every 100 million (Lynch et al., 2001; Lynch and Conery 2000) to 250 million years (McLysaght et al., 2002; Dehal and Boore 2005).

Using the most liberal of estimates (based on eukaryotes and not the more slowly evolving prokaryotes), these 10 duplicative events, beginning with the first gene, would have required from 1 billion to 2.5 billion years, before a minimal gene set of 382 genes could have been established. Since there is evidence of life on this planet, dated from 4.4 to 3.8 bya, and if we accept the latter date and the most liberal genomic duplication rates, this means that the genes necessary for life must have begun to evolve prior to the creation of this planet; at least 4.8 to 5.3 bya. Unless we wish to believe a minimal gene set of 382 life-sustaining genes were created ex-nihilo, or that the first gene underwent an accelerated rate of duplication only to dramatically slow down after life was established, then the only other alternative is the first gene was established prior to the creation of Earth.

5. The First Gene: Time Frame for Genetic & Cellular Evolution

Since genetics and biochemical analysis indicates life was present on this planet in the period between 4.4 bya to 3.8 bya, it can be deduced that these life forms consisted of single celled microbes. However, based on genomics, single celled microbes did not evolve into multi-cellular microbes until nearly 2 billion years had passed.

Evidence based on molecular clocks indicates that multi-cellular eukaryotes with 2 cell types were present by 2.3 bya (Hedges et al. 2004). This data is supported by a genomic analysis of mitochondria, which indicates they took up residence as a distinct entity within eukaryotic cells between 2.3 to 1.8 bya (Hedges et al. 2004; Mentel and Martin 2008); genetic evolutionary transitions which have been directly attributed to alterations in the biosphere and increases in oxygen (Joseph 2000, 2010b). Thus, if genetic-based life appeared on Earth 4.4, 4.2, or 3.8 bya then we can deduce that between 1.5 to 2 billion years passed before the evolution of multicellularity. Further genetic analysis indicates it took another 800 million years to expand from 2 cell types to 10 cell types (Hedges et al. 2004); changes also associated with a changing environment and horizontal gene exchange

(Joseph 2010ab). The eukaryotic genome also increased in size and complexity.

If this data is accepted, then we can also arrive at certain conclusions about the origin of the first gene. We have already determined that it would take 9 to 10 duplicative events to evolve from a single gene to a minimal gene set which can maintain life. Using the most liberal of estimates (based on genomic duplications in the faster evolving eukaryotes), these duplicative events would have required at least 1 billion to 2.5 billion years in order to reach a minimal gene set of 382 genes to 485 genes.

Assuming it takes at least 1.5 billion to 2 billion years (by) to evolve from a single celled microbe to a multi-cell of 2 cell types in response to a changing environment, then it is reasonable to assume it may take at least 1.5 by to evolve from the simplicity of a single celled prokaryote to the first super kingdoms, archea and bacteria. Likewise, it can be estimated that it would take at least 1.5 by to evolve from proto-cell to prokaryote. However, might it take even longer in an unchanging environment? In fact, based on genomics, it can be concluded that 1.5 billion years is an overly conservative estimate as it is based on evolutionary transitions in eukaryotes which "evolves faster than prokaryotes, with those eukaryotes derived from eubacteria evolving faster than those derived from archaeobacteria" (Hedges et al. 2001).

Nevertheless, if we accept that the first super kingdoms, archea and bacteria were established by 4 bya, this would mean that the first prokaryotes must have evolved at least 1.5 by earlier, i.e. 5.5 bya and were preceded by the first proto-cells 7 bya, which were preceded by the establishment of the first gene. This first gene then acquired the ability to replicate and underwent at least 9 to 10 duplicative events over 1 by to 2.5 by in order to create the minimal gene set necessary to sustain the life of a proto-cell. This provides a genetic birth date of 8 by to 9.5 billion years using the most conservative of estimates. However, if we considered the evolution of viruses as a necessary step in the evolution of life, then another 1.5 by could be added to the total, which yields a birth date ranging from 9.5 to 11 bya.

6. Viruses, RNA-Worlds & Replication

Gene and genome duplication requires the ability to replicate, which is made possible by the interactions of DNA and RNA. A number of scientists, beginning with Joseph (2000) have argued that viruses may have inserted DNA and/or RNA into the genomes of the first cells thereby providing the ability to replicate. In this model, first proposed by Joseph (2000) viruses may have served as mobile RNA-worlds (Joseph 2000, 2009, 2010a).

However, even if viruses inserted the necessary replication machinery into the first genes or genomes, it would have still required 9 to 10 duplicative events and from 1 billion to 2.5 billion years to achieve a minimal gene set to maintain life.

The question then becomes: Where did viruses obtain their genes?

One possibility is that viral RNA evolved prior to DNA leading to genes then proto-cells, or that genes evolved first, leading to viruses (mobile RNA Worlds), then proto-cells, then prokaryotes then archae and bacteria and then eukaryotes. If correct, this would push the ancestry of the first gene and genetic-life even further back in time, i.e. from 9.5 to 11 bya using the most conservative of estimates.

7. Statistical Analysis of the Genetics of Evolutionary Change

There are two models of eukaryotic genesis, "genetic merger" and "deep roots." However, if eukaryotes evolved at the same time as archae and bacteria, between 4.4 and 3.8 bya, or at a late date, a statistical analysis of both evolutionary models, and a wide range of dates, still support an extra-terrestrial origin for the genetic basis of life, as does an analysis of genomic evolution in eukaryotes.

Whilst establishing evolutionary connections between species is relatively easy with gene sequencing techniques, setting an absolute timescale for evolution is fraught with problems due to gene duplication and deletion, exon shuffling, and other variables such as the insertion of viral genes. There there is the problem of horizontal gene transfer, such that genomes increase in size due to the acquisition of additional genes inserted by prokaryotes and viruses over billions of years of time. Nevertheless, increases of genome length and complexity as a function of duplicative events, can be assumed to represent a time sequence of evolution, assuming gene loss is equal and provided correction is made for those species that have accumulated larger number of silent (or "non-coding" genes) by gene duplication. Thus, for the purposes of this analysis, only the length of the coding DNA was reckoned as a measure of time elapsed in evolution.

Consider, for example, Figure 1, which shows a relation between the length of the coding genome and the total genome size for a number of cases (Lynch, 2007). We note that for plants and animals a reduction by a factor 10-100 is common.

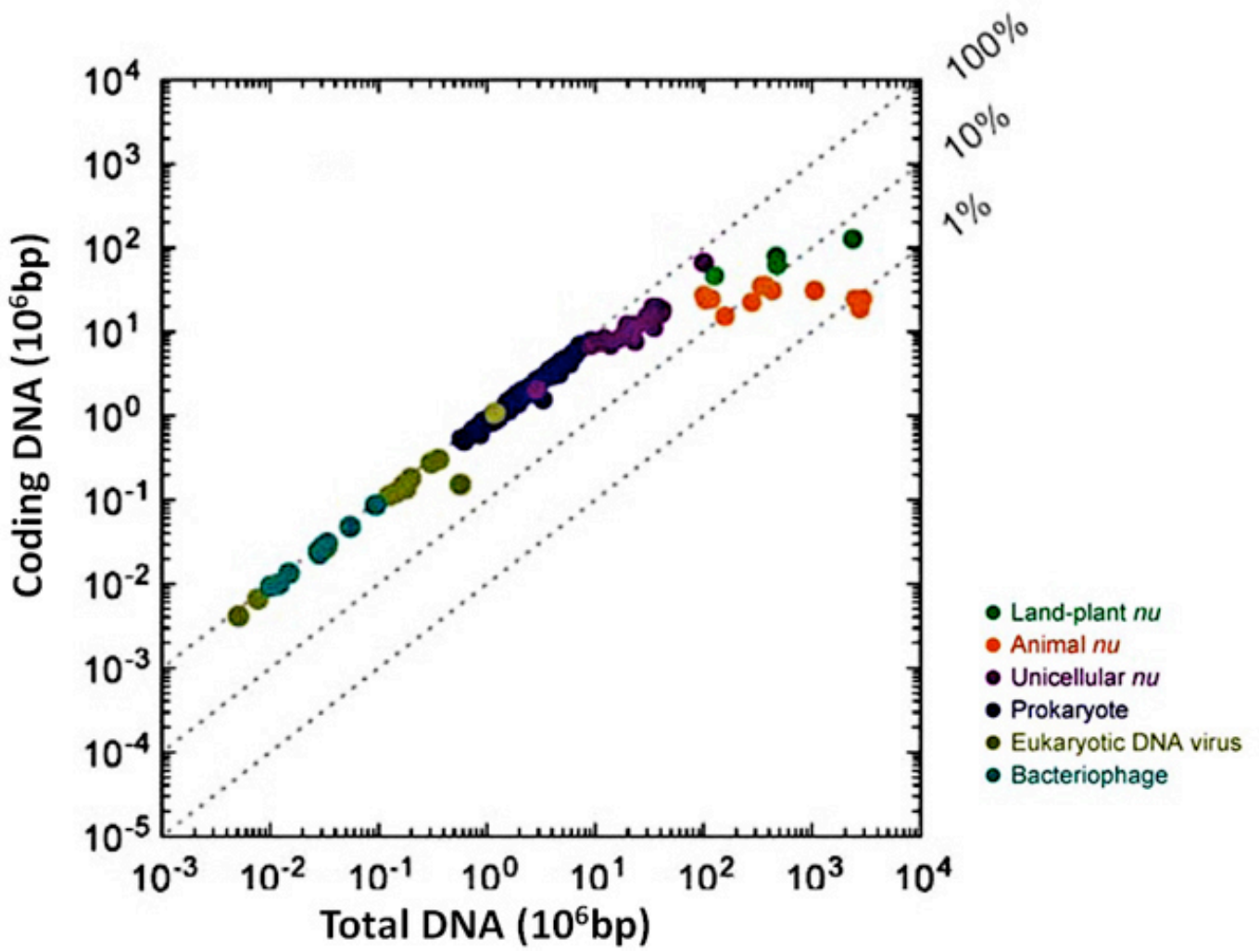


Figure 1. Correlation between coding DNA and total DNA in various species

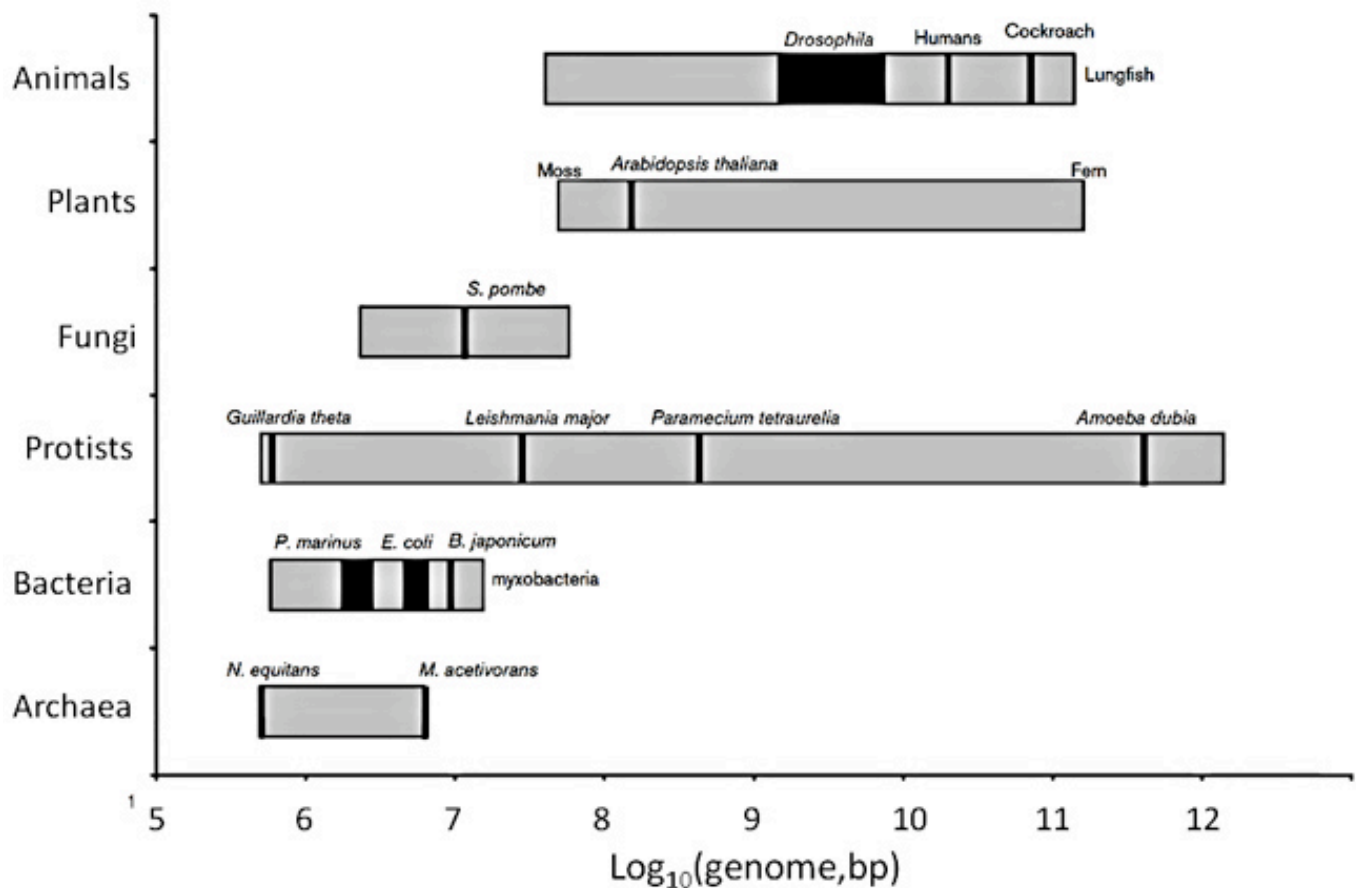


Figure 2. Estimates of genome sizes (Ussery and Hallin, 2004).

In Figure 2 we see that the average genome length increases from archaea to plants and animals with the exception of protists that include fungi, where in some instances the total genome length exceeds that in humans. Even within the animal kingdom, the cockroach and the lungfish are seen to have total genome lengths that exceeds that of humans – gene doublings and non-coding DNA accounting for their extended lengths.

A more reliable measure of the evolution in complexity and increases in the number of genes could be given by the total number of genes coding for proteins. Table 1 gives available data for a representative sample of life forms in various kingdoms and genera.

Table 1

Species	Number of genes	Time of appearance (bya)	
		(A)	(B)
<i>Mycoplasma genitalis</i> (primitive archaee)	382	4.2	
<i>Guillardia theta</i> <i>Crypomonad</i> (Primitive Eukayote)	464	4.2	
<i>Rickettsia prowazekii</i> (First bacterium)	1111	3.8	3.8
<i>Aquifix aeolicus</i> (modern bacterium)	1749	3.5	3.5
<i>Saccarintces cerevisiae</i> (yeast)	6,000	1.8	1.8
<i>Caenorhabditis elegans</i> (roundworm)	19,000	0.54	0.54
<i>Arabidopsis thaliana</i> (flowering plant)	25,000	0.5	0.6
<i>Mus musculus</i> (mouse - mammal)	25,000	0.1	0.1

Although major uncertainties still remain, a timeframe for the evolution life on Earth is available from the geological record, with different species appearing at different times.

Evidence for life on the Earth has been inferred from data on $^{12}\text{C}/^{13}\text{C}$ ratios of carbonaceous material in ancient rocks. Since biology takes up ^{12}C preferentially, an excess of this isotope can be taken as an index of biology. As already noted, evidence of a ^{12}C excess is found in sediments dated at 3.8 bya at the end of the Hadean epoch (Mojzsis et al, 1996) - a time when the Earth was subject to intense bombardment by comets and other extra-terrestrial debris. Even earlier at 4.2bya there is arguable evidence of life, also based on a ^{12}C excess (Nemchin et al. 2008; O'Neil et al. 2008). The evidence of a ^{12}C excess between 4.2bya and 3.8 bya can be most plausibly interpreted as a residue of biology that establish itself despite the hostile conditions prevailing on Earth during the Hadean epoch. In view of the phylogenetic evidence of extensive gene swapping (reviewed in Joseph 2010b) and as there is no evidence that proto-cells ever existed on Earth, it is likely that archaee and bacteria (and perhaps even single celled eukaryotes) were present between 4.2 to 3.8 bya, and exchanging genes. And this leads us again to the question: When did the first genes evolve?

For the purpose of our present argument let us take the most conservative estimates as to the origin of life at around 3.8 bya. As representative points further up in the evolutionary progression we take the appearance of multicellular eukaryotes as at 2.3bya, worms at

1bya, teleosts at 0.5bya and mammals at 0.1 bya .

For the purposes of this statistical analysis the total numbers of genes in the several species listed in Table 1 were plotted against time in Fig.3. The line is a regression line

$$y = 0.3141x + 4.447$$

giving a best fit to the points. Extrapolation to zero on the y axis gives a time of $t=14$ bya. Within the uncertainties of the data plotted in Fig.3 we estimate that this intercept could lie in the range $10.5 < 14.5$ bya.

Even if we ignore the evidence suggestive of biology in this planet's oldest rocks dated to 4.2bya and the genetic evidence indicating the presence of prokaryotes before 4 bya, and accept the 3.8 by date for prokaryotes and not eukaryotes (which according to the "merger model" are assumed to have evolved much later), this still leads to a genetic origin of over 10 bya.

8. Discussion

Genomic analysis, based on increases in genome size and the evolutionary record, indicate that genes began to evolve and to undergo duplicative events billions of years before the formation of this planet, at least 10 billion years ago. This does not mean that life began 10 billion years ago, but rather that the first gene was fashioned approximately 6 billion years before the creation of Earth. The genetic evidence supports extra-terrestrial abiogenesis.

Within the range of validity of the assumptions we have made an exponential dependence of gene number with time which can be inferred over the time range for which data is available. The slope of the regression line is determined mainly by the points for a bacterium, metazoan, and a mammal.

Extrapolation of the regression line backwards in time implies that similar rates of genome and pre-genomic evolution apply prior to the event or events that led to the origin of fully-fledged bacterial and eukaryotic life. A genetic birthdate of between 10.5 to 14.5 billion years, supports the likelihood that the evolution of the entire ensemble of organisms or genes at the base of Woese's tree of life is derived from sources external to Earth and which originated at about the time of the formation of this galaxy.

After the first gene was established, this was followed or was accompanied by the acquisition of the capacity to replicate, until a minimal gene set was established making

life possible.

9. Correlation of Gene Numbers with Time

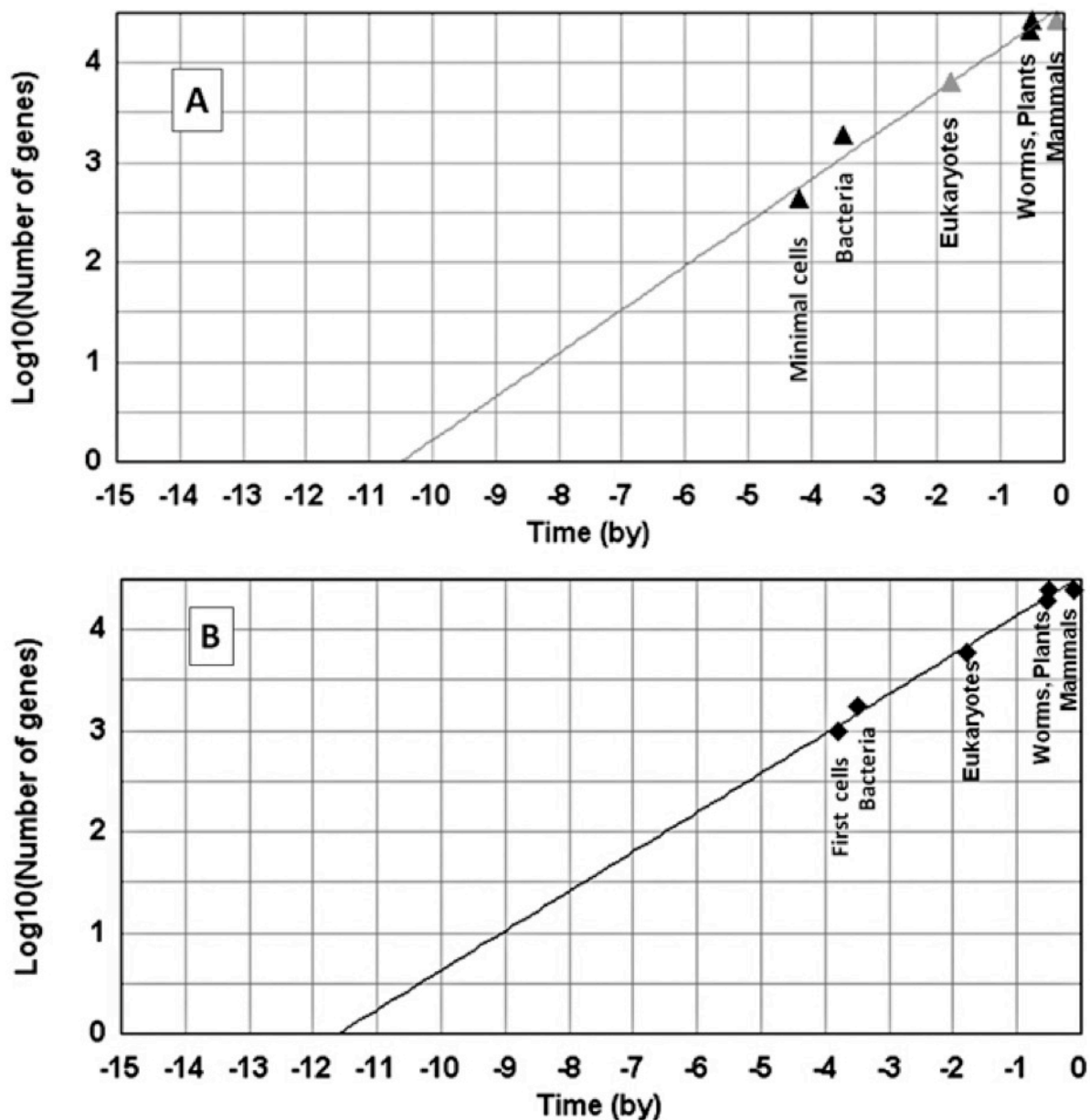
We have noted earlier that microbial life may have been present on Earth between 4.2 and 3.8 bya during the Hadean epoch, and such life may well have included archaea, bacteria as well as eukarya. It can be assumed they must have possessed a genome sufficiently large to make life possible. At a minimum level the bacterial genome must have consisted of at least ~382 genes (equal to a stripped down *Mycoplasma genitalis*), and the first eukaryotes must have also possessed a genome at least the size of the genome of the simplest of eukaryotes, i.e. an enslaved algae, *Guillardia theta Cryomonad* (464 genes). However, it must be emphasized that these are very conservative estimates, and that the original gene ensemble that made life possible may have consisted of hundreds of additional genes, in which case, the ancestry of genetic life would be pushed even further back in time.

Whether such minimal gene numbers increased between 4.2bya to 3.8bya is unknown, and whether or not life was even present 4.2bya is a matter of debate. In contrast, it is widely accepted that life had become well established on Earth by 3.5bya as evidenced by the discovery of microbial fossils resembling cyanobacteria (Schopf et al, 2002, 2007). Therefore, for the purposes of the next statistical analysis we shall assume life was certainly present by 3.5bya, and take this as the possible start date for the onset of the evolution and diversification of species. However this does not mean to imply that the common ancestor for all life can be dated to 3.5bya. For this reason we explore earlier evolutionary steps represented by the columns A, B, C as presenting alternative possibilities.

A stem group of eukaryotes (protocists) have been confirmed to occur in rocks dated 1.8-1.3bya (Knoll et al, 2006). Based on genetic molecular clocks it has also been determined that bilateria and the "lower metazons" and the so called "higher" metazoans, diverged from a common ancestor that lived anywhere from 1.3bya to 830 mya (e.g., Wray et al., 1996; Peterson et al., 2004, Nei et al., 2001; Gu 1998). Based on the fossil record the metazoan explosion of life had its onset around 540 mya. The steps in an evolutionary ladder that we shall proceed to analyse are set out in Table 1.

We next consider the relation between gene numbers (N) and the time of appearance on the Earth (t). We consider three cases: (A) for which the initial colonization of Earth by minimal-sized bacteria and single celled eukaryotes occurred at $t = -4.2\text{by}$; and (B) for

which the initial colonization is by bacteria with an average gene number of ~ 1000 (e.g. *Rickettsia prowazekii*) at $t = -3.8$ by. Case (B) corresponds to the root of the “tree” of terrestrial life starting with simple bacteria/eukarya deposited at the end of the Hadean impact era, and case (C) including both simple bacteria and eukarya (eg. *Encephalitozoon cuniculi*) at the end of the Hadean epoch. The relevant gene numbers for each of these cases along with linear regression lines are plotted against time in Fig. 3 (A,B,C). Within the range of validity of our assumptions the fits to the data are excellent and an exponential dependence of gene number with time can be inferred over the time range for which data is available. The correlation coefficients calculated to be in the range $R^2=0.91-0.99$ give a high level of confidence in extrapolating to $N \sim 1$.



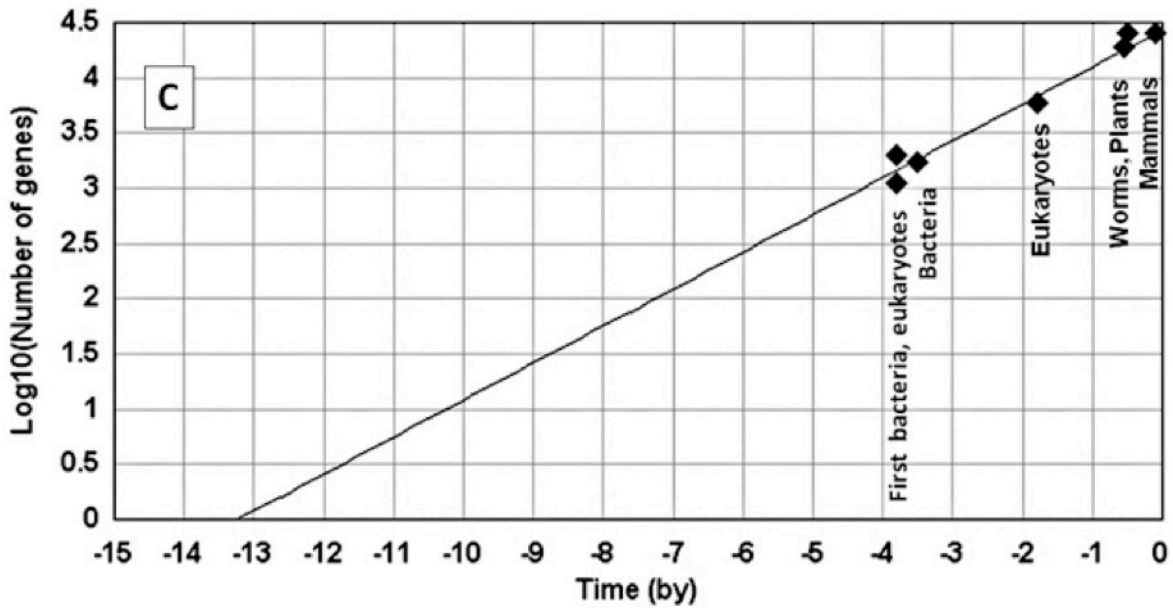


Figure 3. (A,B,C) Correlation between gene numbers and time of appearance of Earth.

Extrapolation to $\log N=0$ on the y axis gives a time estimate for the first origin of genes, and therefore presumably the origin of life. Within the uncertainties of the data plotted in Fig. 3 we estimate that this intercept could lie in the range $10.4 < 13.2$ bya. This data is remarkably consistent with estimates of genetic and cellular evolution detailed in previous and following sections. Thus, the overall pattern of results support the possibility that the first genes may have evolved over 10 bya.

10. Statistical Analysis Based on Gradualism & Evolution of Genome Complexity Also Demonstrates the First Genes Did Not Originate on Earth

Functional complexity of organisms can be approximately measured by the length of the nonredundant functional portion of the genome (Adami et al. 2000). The data on the size of nonredundant functional genome of major phylogenetic lineages was plotted against the time of their origin (Sharov 2006) (Fig. 4). Mammals (on average) have a genome of ca. 3.2×10^9 bp, however only 5% of it is conserved between species (Waterson et al. 2002). Besides conserved regions, there may be additional functional regulatory regions in the genome that are species-specific. These regions, which can be identified based on the absence of transposons, account for 12-20% genome size (Simons et al. 2006). If we take 15% as a rough estimate, then the size of functional and non-redundant genome in

mammals (on average) is 4.8×10^8 bp. Teleosts originated 0.5 billion years ago (Miller et al. 2003), and the fish genome size (on average) is 4×10^8 bp with 1/3 of it occupied by gene loci (Aparicio et al. 2002). Worms may have first appeared 1 billion years ago (Seilacher et al. 1998). The genome size of *Caenorhabditis elegans* is 9.7×10^7 bp and 75% of its length is functional (C.elegans Sequencing Consortium 1998). The smallest eukaryote genome (2.9×10^6 bp) was found in *Encephalitozoon cuniculi* (Katinka et al. 2001), and the smallest prokaryote genome size (5×10^5 bp) was found in *Nanoarchaeum equitans* (Waters et al. 2003) and *Mycoplasma genitalium* (Fraser et al. 1995). Prokaryotes and eukaryotes with the smallest genome are parasitic and may have a reduced genome size due to parasitism. However they were selected them to get the most conservative estimate for the time elapsed since the genetic origin of life.

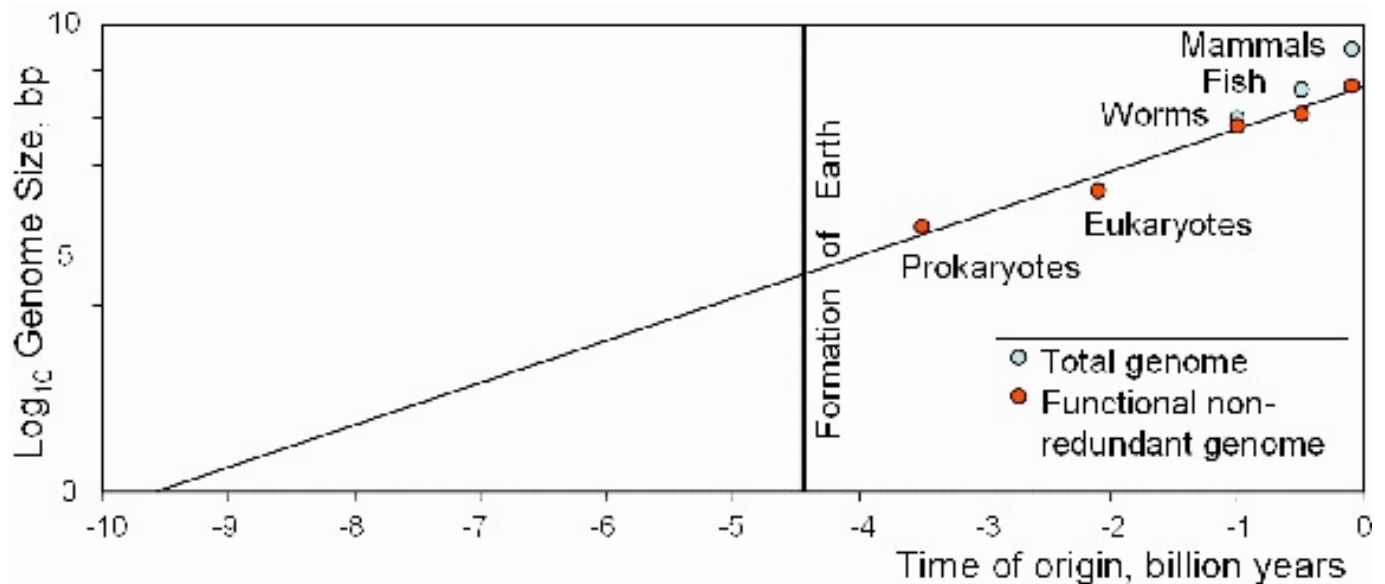


Figure 4. Regression of log genome complexity versus time of origin, modified from (Sharov 2006).

Using the regression of logarithm of functional complexity versus time (Fig. 4), it can be shown that genetic complexity increased exponentially with time, growing slowly ca. 7.8 fold per one billion years (Sharov 2006). This pattern of increase in complexity is consistent with the principle of gradualism. The exponential increase can be explained by several positive feedback mechanisms, which include gene cooperation, gene duplication, and creation of new functional niches for emerging genes (Sharov 2006).

The exponential model of increasing genome complexity can be used to predict the time of the genetic origin of life. Based on the model of life's evolution, beginning from a single coding elements, the data indicates that the genetic basis of life originated at the time point where log genome complexity was zero. Based on the regression of log genome complexity versus time, the origin of the genetic basis of life is projected around 10 billion years ago (Fig. 4). Because two earliest points on the graph are most uncertain, Sharov (2006) did a sensitivity analysis by varying these points within the limits of uncertainty (± 300 Mya, and ± 0.3 log bp). With these variations the date of life origin may vary from 7 to 13 billion years which is still greater than the age of Earth and which is consistent with the analysis performed in Sections 8, 9 and 10, and the first half of this paper.

11. Speculation: Did Life Begin Following the Big Bang?

There are a variety of scientific views as to the nature and origin of this universe, the most popular of which is the "Big Bang" which is estimated to have taken place approximately 10 to 14 billion years ago, with data from the WMAP satellite (based on detailed analysis of the cosmic microwave background) providing evidence for a date of 13.7 bya. This later date has been accepted by current consensus. This range of birth dates for this universe is remarkably consistent with the data obtained from two of the genetic analyses reported in this paper, i.e. $10.5 < 14.5$ bya. (Section 7), $10.4 < 13.2$ bya. (Section 8).

To speculate: if there was a Big Bang beginning to this universe, then the data reported in Sections 7 and 8, could be interpreted to mean that the genetic origins of life can also be traced backwards in time to the "Big Bang". If correct, and if there was a Big Bang, then this creation event not only created this universe, but the conditions and essential elements necessary for life in this universe.

Simply put: it could be argued that following Big Bang nucleosynthesis when the universe had cooled sufficiently to create elemental abundances, that the ensuing nucleogenesis and nucleosynthesis led to the production of elements essential for life including carbon (due to triple collisions of helium-4 nuclei) and then the production of molecules such as purine and pyrimidines and the nucleotides adenine, guanine, cytosine, thymine, and uracil, thereby providing all the essential elements for the construction of DNA and RNA. All this *could* have taken place within several hundred millions years after the presumed Big Bang creation of this universe.

To speculate further: it could be said that if the Big Bang cosmological model is correct, then the Big Bang not only created the essential elements necessary for life, but a Universe

in which that life could dwell--a universe ideally fit for incubating life.

12. CONCLUSIONS

Life appeared on this planet between 4.4 to 3.8 billion years ago, with genetic and biological indices indicating the presence of life by at least 4.2 bya. These Earthly-life forms, and most certainly their prokaryotic and/or proto-cellular ancestors, must have possessed a minimal gene set, which made life possible. This minimal gene set must have been established following the same genetic processes of duplication, replication, and gene deletion, as is characteristic of life on this planet.

For example, there is considerable evidence that the entire eukaryotic genome underwent duplication at the onset of eukaryotic evolution (Makarova et al., 2005). These genes then continued to undergo repeated episodes of single gene and whole genome duplication such that the eukaryotic genome increased in size (Kellis et al., 2004; Dietrich et al., 2004; Dehal and Boore 2005) thereby triggering the transition and divergence between numerous species, ranging from yeast and fungi (Liti and Louis, 2005) to chordates and non-chordates (Dehal and Boore 2005; McLysaght et al., 2002). Likewise we can conclude that WGD and increases in gene numbers led to archae and bacteria diverging from prokaryotes, which was preceded by prokaryotes diverging from proto-cells.

Increases in the number of genes and the size of the genome, coupled with estimates as to the length of time that must elapse between whole genome duplications, can provide gross estimates as to the date of origin of the first genes and then a minimal period of time before a minimal number of genes necessary for life were established, i.e. at least 9 to 10 duplicative events which would have required at least 1 billion to 2.5 billion years--which is a very conservative estimate. As based on estimates of whole genome duplicative events and the transitions from single cell to multi-cellularity, these initial duplicative events, beginning with the first gene and leading to a minimal gene set and genome, and then the divergence of prokaryotes from proto-cells, followed by the divergence of prokaryotes to the superkingdoms of archae and bacteria, must have taken place over nearly 10 billions of years of time, beginning well before the establishment of life on this planet (4.4 to 3.8 bya) and thus prior to the creation of Earth.

These assumptions are supported by the genetic analyses reported in this paper, and which demonstrate that the ancestry of genetic life extends backwards in time over 10 billion years to a period early in the evolution of this galaxy, i.e. $10.5 < 14.5$ bya. (Section 7), $10.4 < 13.2$ bya. (Section 8); data which coincides with the Big Bang model for the origin of

the universe. This does not mean that life began over 10 billion years ago, but rather that the first gene was fashioned billions of years before the creation of Earth.

Acknowledgements: We thank the referees, and Dr. Alexei Sharov, Dr. Martin Line, Dr. Ed Trifonov, Dr. Antonio Bianconi and professor Pabulo Henrique Rampelotto for their helpful comments. We also thank Dr. Sharov for granting permission to present his genetic analysis (Section 10) first reported in 2006.

References

- Adami, C., Ofria, C., Collier, T. C.. (2000). Evolution of biological complexity. *Proc. Natl. Acad. Sci. U. S. A.*, 97, 4463-4468.
- Altermann, W. and Schopf, J.W., (1995). Microfossils in Neoproterozoic Cambell group Griqualaul West Sequence of Transvaal Supergroup and evolutionary and other pleoenvironmentsa implications, *Precambrian Res*, 75 (1-2), 65-90.
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. et al. (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297, 1301-1310.
- Aravind, L., Watanabe, H., Lipman, D.J., & Koonin, E.V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl. Acad. Sci.* 97, 11319-11324.
- Bejerano, G., (2004). Ultraconserved Elements in the Human Genome *Science*, 304. 1321 - 1325.
- Blanc, G., Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*. 16, 1667–1678.
- Breitbart M, Rohwer F. (2005). Here a virus, there a virus, everywhere the same virus?. *Trends Microbiol*, 13(6):278–84.
- Cavalier-Smith, T., (2006). Cell evolution and Earth history: stasis and revolution--*Phil. Trans. R. Soc. B* 29, 361, 1470 969-1006.
- C.elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, 282, 2012-2018.

- Charlebois, R.L., & Doolittle, W.F. (2004). Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.* 14, 2469–2477.
- Clark, C. G. (1990). Genome Structure and Evolution of *Naegleria* and its Relatives. *Journal of Eukaryotic Microbiology*, 37, Issue 4, pages 2s–6s.
- Conley, A. B., Piriyaopongsa, J., Jordan, I. K. (2008). Retroviral promoters in the human genome, *Bioinformatics*, 24, 1563-1567.
- Crick, F. (1981). *Life Itself. Its Origin and Nature.* Simon & Schuster, New York.
- Dayhoff MO, Barker WC, McLaughlin PJ. (1974). Inferences from protein and nucleic acid sequences: early molecular evolution, divergence of kingdoms and rates of change. *Orig. Life.* 5:311–330.
- Dayhoff MO, Barker WC, Hunt LT. (1983). Establishing homologies in protein sequences. *Methods Enzymol.* 91:524–545.
- Dehal, P., & Boore, J.L.. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3, 314.
- De Rosa M, Gambacorta A, Gliozzi A (1986). Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. *Microbiol. Rev.* 50 (1): 70–80.
- Dombrowski, H. (1963). Bacteria from Paleozoic salt deposits. *Ann NY Acad Sci* 10⁸, 453 ±460.
- Dombrowski, H. J. (1966). Geological problems in the question of living bacteria in Paleozoic salt deposits. In *Second Symposium on Salt*, vol. 1. 44-78.
- Dose, K. (1988). The origin of life: More questions than answers. *Interdisciplinary Science Review*, 13, 348-356.
- Douzery E.J, Snell E.A, Baptiste E, Delsuc F, Philippe H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl Acad. Sci. USA.* 101:15386–15391.
- Durand, D. (2003). Vertebrate evolution: doubling and shuffling with a full deck. *Trends Genet.* , 19, 2–5.
- Eck RV, Dayhoff MO. (1966). Evolution of the structure of ferredoxin based on living

relics of primitive amino acid sequences. *Science* 152:363–366.

Feng D-F, Cho G, Doolittle RF. (1997) Determining divergence times with a protein clock: update and reevaluation. *Proceedings of the National Academy of Sciences (USA)*. 94:13028–13033.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science*, 270, 397-404.

Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Chapman, J., Pham, J., and 14 more author (2010). The Genome of *Naegleria gruberi* Illuminates Early Eukaryotic Versatility, *Cell*, 140, 631-642.

Furnes, H., Banerjee, N. R., Muehlenbachs, K., Staudigel, H., de Wit, M. (2004). Early life recorded in archean pillow lavas. *Science*, 304, 578-581.

Gibson, C. H., Wickramasinghe, N. C., Schild, R. (2010). First life in the oceans of primordial-planets: The biological big bang. *Journal of Cosmology*, 11. 3490-3489.

Glass, J., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Hutchison, III, C. A., Smith, H.O., and Venter, J. C. (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A*. 2006 January 10; 103(2): 425–430.

Gregory, T.R. (2005). Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics* 6: 699-708.

Gregory, T.R. and DeSalle, R. (2005). Comparative genomics in prokaryotes. In: *The Evolution of the Genome*, edited by T.R. Gregory, pp. 585-675. Elsevier, San Diego, CA.

Gu, X. (1998) Early Metazoan Divergence Was About 830 Million Years Ago. *J. Mol. Evol.* 47, 369-371.

Harris, J. K., et al., (2003). The Genetic Core of the Universal Ancestor Genome *Res.*13, 407-412.

Hartman, H., & Fedorov, A. (2002). *Proceedings of the National Academy of Sciences*, 99, 1420.

Hedges, S. B. (2009). Life. In, *the Timetree of Life*. S. B. Hedges & S. Kumar, Eds.

Oxford University Press.

Hedges SB, Blair JE, Venturi ML, Shoe JL., (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol.* Jan 28;4:2.

Hedges SB, Chen H, Kumar S, Wang DY, Thompson AS, Watanabe H. (2001). A genomic timescale for the origin of eukaryotes. *BMC Evol Biol.* 2001;1:4.

Hoyle, F., (1982), *Evolution from Space (The Omni Lecture)* Enslow Publishers, USA.

Hoyle, F., Wickramasinghe, N.C. (1980). *Evolution from Space.* J..M. Dent & Sons.

Hoyle, F. Wickramasinghe, N. C. (2000). *Astronomical Origins of Life. Steps Towards Panspermia,* Kluwer Academic Publishers.

IHGSC (2001). International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

Jaillon O, et al. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature.* 431:946–957.

Jalasvuori, M., and Jaana K.H. Bamford, (2010). Viruses and Life: Can There Be One Without the Other? *Journal of Cosmology,* Vol 10, 3446-3454.

Jeffreys AJ, et al. (1980). Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization. *Cell.* 21:555–564.

José, M. V., Morgado, E. R., Govezensky, T. (2010). How universal is the universal genetic code? A question of extraterrestrial origins. *Journal of Cosmology,* 5, 854-874.

Joseph, R. (2000). *Astrobiology, the death of Darwinism and the origins of life.* University Press. California.

Joseph, R. (2009). Life on Earth came from other planets. *Journal of Cosmology,* 1, 1-56.

Joseph, R. (2010a). The origin of eukaryotes: Archae, bacteria, viruses and horizontal gene transfer. *Journal of Cosmology,* 10, 3418-3445.

Joseph (2010b) Climate change: The first four billion years. The biological cosmology of global warming and global freezing. *Journal of Cosmology,* 2010, 8, 2000-2020.

Joseph, R., Schild, R. (2010). *Biological cosmology and the origins of life in the universe.*

Journal of Cosmology, 5, 1040-1090.

Karner MB, DeLong EF, Karl DM (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409 (6819): 507–510.

Katinka, M.D., et al. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, 414, 450–453.

Kellis M, Birren BW, Lander ES. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*. 2004;428:617–624.

Kimura, H., J.-I. Ishibashi, H. Masuda, K. Kato, and S. Hanada (2007). Selective Phylogenetic Analysis Targeting 16S rRNA Genes of Hyperthermophilic Archaea in the Deep-Subsurface Hot Biosphere *Appl. Environ. Microbiol.* 73:2110-2117.

Kimura, H., M. Sugihara, K. Kato, and S. Hanada (2006). Selective Phylogenetic Analysis Targeted at 16S rRNA Genes of Thermophiles and Hyperthermophiles in Deep-Subsurface Geothermal Environments *Appl. Environ. Microbiol.* 72:21-27.

Knoll AH et al., (2006). Eukaryotic organisms in Proterozoic oceans. *Philos Trans R Soc Lond B Biol Sci.* 36, 1023-1038.

Koonin, E.V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Rev. Microbiol.* 1, 127–136.

Koonin EV. (2009) Evolution of genome architecture. *Int. J. Biochem. Cell Biol.* 41:298–306.

Koonin, E.V., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5, R7.

Koonin, E. V., & Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. *Trends Genet*, 21, 647-654.

Koonin, E.V., & Wolf, Y.I. (2008). Genomics of bacteria and archaea: the emerging generalizations after 13 years. *Nucleic Acids Res.* 36, 6688–6719.

Kunin V, Ouzounis CA. (2003) The balance of driving forces during genome evolution in prokaryotes. *Genome Res.* 13:1589–1594.

- Kuppers, B.O. (1990). *Information and the origin of life*. Cambridge, MA: MIT Press.
- Kurland, C. G., Collins, L. J., & Penny, D. (2007). *Science*, 316, 543.
- Leininger S, Urich T, Schloter M, et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442 (7104): 806–809.
- Line, M.A. (2010). Extraterrestrial Origin of Life and Genetic Gradualism: Commentary *Journal of Cosmology*, 5, 851-853.
- Lynch, M. (2006). Streamlining and simplification of microbial genome architecture, *Annual Review of Microbiology* 60: 327-349.
- Lynch, M. (2007). *The Origins of Genome Architecture*, Sinauer Associates, Sunderland, MA.
- Lynch, M., & Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*. 290, 1151–1155.
- Lynch M, O'Hely M, Walsh B, Force A. The probability of preservation of a newly arisen gene duplicate. *Genetics*. 2001;159:1789–1804.
- Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, Han CS, Lapidus A, Metcalf WW, Saunders E, Tapia R, et al. (2006). The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J. Bacteriol.* 188:7922–7931.
- Makarova, K. S., et al., (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell *Nucleic Acids Research*, 33(14):4626-4638.
- Manning, C. E., Mojzsis, S. J., Harrison, T. M. (2006). Geology, age and origin, of supracrustal rocks at Akilia, West Greenland. *American Journal of Science*, 306, 303-366.
- Margulis, L., Sagan, D., & Thomas, L. (1997). *Microcosmos; Four Billion Years of Evolution from Our Microbial Ancestors*, University of California Press.
- Martin, W., & Muller, M. (1988). The hydrogen hypothesis for the first eukaryote. *Nature*, 392, 37–40.
- Mastrapaa, R.M.E., Glanzberg, H., Head, J.N., Melosha, H.J, Nicholson, W.L. 2001. Survival of bacteria exposed to extreme acceleration: implications for panspermia, *Earth*

and Planetary Science Letters 189, 30 1-8.

McGenity, T. J., Gemmell, R. T., Grant, W. D. & Stan-Lotter, H. (2000). Origins of halophilic microorganisms in ancient salt deposits. *Environ Microbiol* 2, 243-250.

McLean, R. J. C., and McLean, M. A. C. (2010). Microbial survival mechanisms and the interplanetary transfer of life through space. *Journal of Cosmology*, 7, 1802-1820.

McLysaght, A., Hokamp, K., & Wolfe, K.H., (2002). Extensive genomic duplication during early chordate evolution. *Nat Genet.* 31, 282-9.

Medstrand P, van de Lagemaat LN, Mager DL. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12:1483 -1495.

Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL. (2005). Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res.* 110(1-4):342-52.

Mentel, M. Martin, W. (2008) Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry *Phil. Trans. R. Soc. B* 27 363 no. 1504 2717-2729.

Miller, R. F., Cloutier, R., Turner, S. (2003). The oldest articulated chondrichthyan from the Early Devonian period. *Nature*, 425, 501-504.

Mirkin BG, Fenner TI, Galperin MY, Koonin EV. (2003). Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* 3:2.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P., Friend, C.R.L. (1996). Evidence for life on Earth before 3,800 million years ago. *Nature* 384, 55–59.

Moran, N.A. (2002). Microbial minimalism: Genome reduction in bacterial pathogens. *Cell*, 108 583-586.

Mushegian, A, (2008). Gene content of LUCA, the last universal common ancestor. *Front Biosci.* 13:4657-66.

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M.

- (2006). The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science* 314:267.
- Nei, M., Xu, P. & Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms *Proc. Natl. Acad. Sci. USA* 98, 2497-2502.
- Nemchin, A. A., Whitehouse, M.J., Menneken, M., Geisler, T., Pidgeon, R.T., Wilde, S. A. (2008). A light carbon reservoir recorded in zircon-hosted diamond from the Jack Hills. *Nature* 454, 92-95.
- Nickerson, C.A., Ott, C.M., Wilson, J.W., Ramamurthy, R., Pierson, D.L., (2004). Microbial responses to microgravity and other low-shear environments. *Microbiology and Molecular Biology Reviews* 68, 345-361.
- Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., Setlow, P. (2000). Resistance of Bacillus Endospores to Extreme Terrestrial and Extraterrestrial Environments, *Microbiology and Molecular Biology Reviews* 64, 548-572.
- O'Neil, J., Carlson, R. W., Francis, E., Stevenson, R. K. (2008). Neodymium-142 Evidence for Hadean Mafic Crust *Science* 321, 1828 - 1831.
- Peterson, K. J., et al., (2004). Estimating metazoan divergence times with a molecular clock -*PNAS*, 101, 6536-6541.
- Pflug, H. D. (1978). Yeast-like microfossils detected in oldest sediments of the earth *Journal Naturwissenschaften* 65, 121-134.
- Poccia, N., Ricci, A., Bianconi, A. (2010). The emergence of life in the universe at the epoch of dark energy domination. *Journal of Cosmology*, 5. 875-882.
- Poole, A. M., & Penny, D. (2007). *Bioassays*, 29, 74.
- Rivera, M.C., & Lake, J.A. (2004). The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*, 431, 152–155.
- Robertson C, Harris J, Spear J, Pace N (2005). Phylogenetic diversity and ecology of environmental Archaea. *Curr Opin Microbiol* 8 (6): 638–642.
- Rosing, M. T. (1999). C-13-depleted carbon microparticles in > 3700-Ma sea-floor

sedimentary rocks from west Greenland. *Science* 283, 674-676.

Rosing, M. T., Frei, R. (2004). U-rich Archaean sea-floor sediments from Greenland - indications of > 3700 Ma oxygenic photosynthesis. *Earth and Planetary Science Letters* 217, 237-244.

Scannell, D.R., Butler, G., & Wolfe, K.H. (2007) Yeast genome evolution—the origin of the species. *Yeast*. 24, 929–942.

Schoenberg, R., Kamber, B.S., Collerson, K.D., Moorbath, S. (2002). Tungsten isotope evidence from approximately 3.8-Gyr metamorphosed sediments for early meteorite bombardment of the Earth. *Nature* 418, 403–405.

Schmidt A, Ragazzi E, Coppellotti O, Roghi G (2006). A microworld in Triassic amber. *Nature* 444 (7121): 835.

Schneiker S et al. (2007). "Complete genome sequence of the myxobacterium *Sorangium cellulosum*". *Nature Biotechnology* 25 (11): 1281–1289.

Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, Bartels D, Bekel T, Beyer S, Bode E, et al. (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* 25:1281–1289.

Schopf, J.W. et al, (2002). Laser-Raman imagery of Earth's earliest fossils, *Nature*, 416, 73-76.

Schopf, J. W., A. B. Kudryavtsev, A. D. Czaja, and A. B. Tripathi. (2007). Evidence of Archean life: Stromatolites and microfossils: *Precambrian Research* 158:141–155.

Seilacher, A., Bose, P. K., Pfluger, F. (1998). Triploblastic animals more than 1 billion years ago: trace fossil evidence from India. *Science*, 282, 80-83.

Sharov, A.A. (2006). Genome increase as a clock for the origin and evolution of life, *Biol Direct*, 1, 17/

Sharov, A.A. (2009). Exponential increase of genetic complexity supports extra-terrestrial origin of life, *Journal of Cosmology*, 2009, 1, 63-65.

Sharov, A.A., (2010). Genetic gradualism and the extra terrestrial origin of life, *Journal of Cosmology*, 5, 833-842.

- Simons, C., Pheasant, M., Makunin, I. V., Mattick, J. S. (2006). Transposon-free regions in mammalian genomes. *Genome. Res.*, 16, 164-172.
- Snel B, Bork P, Huynen MA. (2002) Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* 12:17–25.
- Soltis, D.E., Bell, C.D., Kim, S., & Soltis, P.S. (2008). Origin and early evolution of angiosperms. *Ann. N Y Acad. Sci.* 1133, 3–25.
- Stan-Lotter, H. et al. (2002). *Halococcus dombrowskii* sp. nov., an archaeal isolate from a Permian alpine salt deposit. *International Journal of Systematic and Evolutionary Microbiology* (2002), 52, 1807–1814.
- Sullivan, N.J, Geisbert TW, Geisbert JB, Shedlock DJ, Xu L, et al. (2006) Immune Protection of Nonhuman Primates against Ebola Virus with Single Low-Dose Adenovirus Vectors Encoding Modified GPs. *PLoS Med* 3(6): e177.
- Taylor JS, et al. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci.* 356:1661–1679.
- Tymowska J, Fischberg M, Tinsley RC. (1977). The karyotype of the tetraploid species *Xenopus vestitus* Laurent (Anura: pipidae). *Cytogenet Cell Genet.* 19:344–354.
- Ussery, D.W., Hallin, P.F. (2004). *Microbiology Comment*, 150, 513-516.
- Van de Peer Y, et al., (2003). Are all fishes ancient polyploids? *J Struct Funct Genomics.* 3:65–73.
- Vision TJ, Brown DG, Tanksley SD. (2000). The origins of genomic duplications in *Arabidopsis*. *Science.* 290:2114–2117.
- Vreeland, R.H., Rosenzweig, W.D., Powers, D.W., (2000), Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature*, 407, 897-900.
- Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., et al. (2003). The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U. S. A.*, 100, 12984-12988.
- Waterston, R. H., Lindblad-Toh. K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing

and comparative analysis of the mouse genome. *Nature*, 420, 520-562.

Wickramasinghe, J., Wickramasinghe, N. C., Napier, W. (2010). *Comets and the Origin of Life*, World Scientific Publ.

Williamson, S. J., et al., (2008). The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples. *PLoS ONE*. 2008; 3(1): e1456.

Woese C, Kandler O, Wheelis M. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, *Proc Natl Acad Sci USA* 87 (12): 4576–9.

Woese CR (2004). A new biology for a new century. *Microbiol. Mol. Biol. Rev.* 68 (2): 173–86.

Wolfe, K.H., & Shields, D.C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387, 708–713.

Wong S, Butler G, Wolfe KH. (2002). Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc Natl Acad Sci U S A*. 99:9272–9277.

Wray, G. A., Levinton, J. S., Shapiro, L. H., (1996). Molecular Evidence for Deep Precambrian Divergences Among Metazoan Phyla- *Science*, 274. 568 - 573.

Yockey, H.P. (1977). A calculation of the probability of spontaneous biogenesis by information theory. *Journal of Theoretical Biology*, 67, 377-398.