ASTROBIOLOGY
Volume 2, Number 3, 2002

Does the Rapid Appearance of Life on Earth Suggest that Life Is Common in the Universe?

CHARLES H. LINEWEAVER[1,2] and TAMARA M. DAVIS[1]

ABSTRACT

It is sometimes assumed that the rapidity of biogenesis on Earth suggests that life is common in the Universe. Here we critically examine the assumptions inherent in this if-life-evolved-rapidly-life-must-be-common argument. We use the observational constraints on the rapidity of biogenesis on Earth to infer the probability of biogenesis on terrestrial planets with the same unknown probability of biogenesis as the Earth. We find that on such planets, older than $\sim$1 Gyr, the probability of biogenesis is $>$13% at the 95% confidence level. This quantifies an important term in the Drake Equation but does not necessarily mean that life is common in the Universe.
Key Words: Biogenesis—Drake Equation. Astrobiology 2, 293–304.

THE BIOGENESIS LOTTERY

MUCH OF CURRENT ASTROBIOLOGICAL RESEARCH is focused on learning more about the early evolution of the Earth and about the origin of life. We may be able to extrapolate and generalize our knowledge of how life formed here to how it might have formed elsewhere. Indirect evidence suggesting that life may be common in the Universe includes:

- Sun-like stars are common.
- Formation of Earth-like planets in habitable zones around these stars may be a common feature of star formation (Kasting *et al.,* 1993; Wetherill, 1996; Lissauer and Lin, 2000; Lineweaver, 2001).
- Life's chemical ingredients—water, amino acids, and other organic molecules—are common (Cronin, 1989; Trimble, 1997; Charnley *et al.,* 2002).
- Sources of free energy such as starlight and reduction–oxidation pairs are common (Nealson and Conrad, 1999).

It is difficult to translate this circumstantial evidence into an estimate of how common life is in the Universe. Without definitive detections of extraterrestrial life we can say very little about how common it is or even whether it exists. Our existence on Earth can tell us little about how common life is in the Universe or about the probability of biogenesis on a terrestrial planet because, even if this probability were infinitesimally small and there were only one life-harboring planet in the Universe, we would, of necessity, find ourselves on that planet. However, the rapidity with which life appeared on Earth gives us more information. If life were rare it would be unlikely that biogenesis would occur as rapidly as it seems to have occurred on Earth.

Although we do not understand the details of how life originated, we have some useful observational constraints on how long it took. Carbon

---

[1] School of Physics, University of New South Wales, Sydney, Australia.
[2] Australian Centre for Astrobiology, Macquarie University, Sydney, Australia.

isotopic evidence suggests that life existed on Earth >3.85 billion years ago (Mojzsis *et al.,* 1996). High temperatures and large frequent sterilizing impacts may have frustrated an earlier appearance of life (Maher and Stevenson, 1988; Sleep *et al.,* 1989). If life originated on Earth, then increasingly tight observational constraints indicate that biogenesis was rapid (Oberbeck and Fogleman, 1989; Sleep *et al.,* 2001). The extraterrestrial implications of rapid biogenesis on Earth and the extent to which this rapidity suggests that life is common in the Universe have not been looked at carefully and are the focus of this paper.

The basic concept is simple: Over a given time period, more probable events happen more often (and thus more rapidly) than less probable events. Thus, the probability of winning a lottery can be inferred from how quickly a lottery winner has won. For example, suppose we have no idea about the probability q, of winning a daily lottery ($0 \leq q \leq 1$). Suppose a gambler buys a lottery ticket every day for 3 days, losing on the first 2 days and winning on the third. We can use this information to infer something about q. Specifically, in this case, we can say that q is more likely to be about one-third than one-hundredth, and is unlikely to be close to 1 [see $\mathscr{L}(n = 3; q)$ in Fig. 1]. If the gambler can only tell us that he won at least once within 3 days, then we can no longer exclude high values of q with such confidence, but the likelihood of q can still tell us that $q > 0.16$ at the 95% confidence level [see $\mathscr{L}(\leq 3; q)$ in Fig. 1 and Eq. A2 of Appendix].

Suppose there is a group of gamblers, all of whom have won at least once within N (e.g., 12) days. A gambler is chosen from this group at random, and after carefully examining his tickets, he tells us that he won at least once within the first 3 days—relatively early in the 12 days that he had to have won by, to be in the group. This is analogous to our situation on Earth. We find ourselves in the group of planets on which biogenesis has necessarily occurred; we have of necessity won the biogenetic lottery some time in the past. And we also find that biogenesis has occurred rapidly. We won soon after life became possible on Earth. Given the above information about the gambler, the likelihood of q is plotted in Fig. 1 as $\mathscr{L}(n \leq 3, N \leq 12; q)$ and allows us to conclude that $q > 0.12$ at the 95% confidence level. This statistical result applies to a group of gamblers or to a group of terrestrial planets on which the probability q of biogenesis is unknown, as long as q is approximately the same for each planet in the group and approximately the same as it was on Earth. In the next section we review the observational constraints on when and how quickly life appeared on Earth. We then use these constraints to identify and critically examine selection effects that complicate this result. Finally, we discuss the relationship between our result, the Drake Equation, and the larger question: "How common is life in the Universe?" Mathematical details are given in the Appendix.

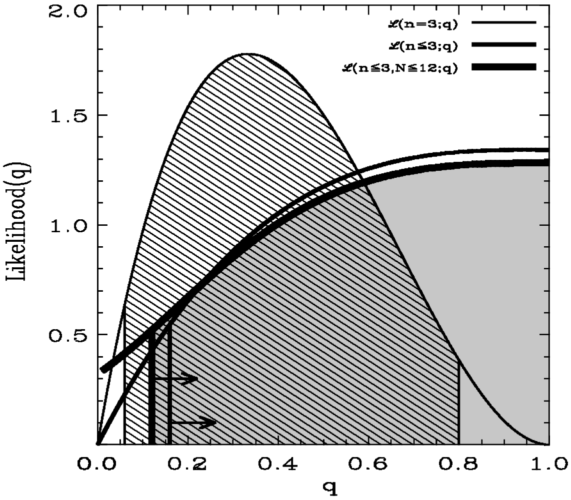

FIG. 1. Let the unknown probability of winning a daily lottery be *q*. Suppose a gambler buys a ticket each day and wins on the third day. From this information we can calculate the likelihood of q [see $\mathscr{L}(n = 3; q)$ above]. The most likely value is, as expected, 1/3, but we can also conclude that $0.06 < q < 0.80$ at the 95% confidence level (hatched area). Different scenario: Suppose that after 3 days the gambler tells us that he won at least once. The likelihood of q then becomes $\mathscr{L}(n \leq 3; q)$ plotted above. High values of q can no longer be excluded, and we can say only that $0.16 < q$ at the 95% confidence level. Different scenario (and one more analogous to our situation on Earth): We have a group of gamblers, all of whom have won at least once, on or before the Nth (e.g., 12th) day. One of them, chosen at random (analogous to the Earth), won at least once, on or before the third day. This is quite early since it could have happened any time during the 12 days. Given this information, the likelihood of q becomes the thick curve labeled "$\mathscr{L}(n \leq 3, N \leq 12; q)$." In this case, we can say that $q > 0.12$ at the 95% confidence level (gray area). See Appendix for computational details.

OBSERVATIONAL CONSTRAINTS ON THE TIMING OF TERRESTRIAL BIOGENESIS

If life originated on Earth, then during and immediately following the Earth's formation there was a period without life ($\Delta t_{\text{frustrated}}$), followed by a period during which life evolved ($\Delta t_{\text{biogenesis}}$), followed by a period during which life has been present (Δt_{life}). The sum of these intervals adds up to the age of the Earth (Fig. 2):

$$\Delta t_{\text{frustrated}} + \Delta t_{\text{biogenesis}} + \Delta t_{\text{life}} = \Delta t_{\text{Earth}} \quad (1)$$

where $\Delta t_{\text{Earth}} = 4.566 \pm 0.002$ Gyr (Allègre *et al.*, 1995). As older fossils and biosignatures have been found, Δt_{life} has gotten longer. The significance of large impacts in frustrating or sterilizing protolife has only recently been appreciated and assessed ($\Delta t_{\text{frustrated}}$). Combined, these observations indicate that biogenesis was rapid since $\Delta t_{\text{biogenesis}}$ is caught in the middle—the longer $\Delta t_{\text{frustrated}}$ and Δt_{life} get, the shorter $\Delta t_{\text{biogenesis}}$ must get. The distinction between how rapid biogenesis was and when it was is important because our result, the inferred probability of biogenesis, depends on how rapid it was, while only a marginal selection effect depends on when it was (see Selection Effects).

The majority of the Earth's mass accreted from planetesimals within the first 100 million years of the Earth's formation (Halliday, 2000). With an initially molten surface, life could not have appeared. The transition from accretion to heavy bombardment included the formation of the Moon by the collision with a Mars-sized object ~4.5 Gyr ago (Hartmann and Davis, 1975; Canup and Asphaug, 2001; Halliday, 2001). We can infer from the dates and sizes of lunar impact craters, whose record goes back to when the Moon formed a solid crust [~4.44 Gyr ago (Sleep *et al.*, 1989)] that the surface of the Earth was periodically vaporized. Since the mass of the Earth is 80 times the mass of the Moon, impacts on the Earth were more numerous and more energetic and periodically produced 2,000K rock vapor atmospheres that lasted for several thousand years (Hartmann *et al.*, 2000; Sleep *et al.*, 2001). These

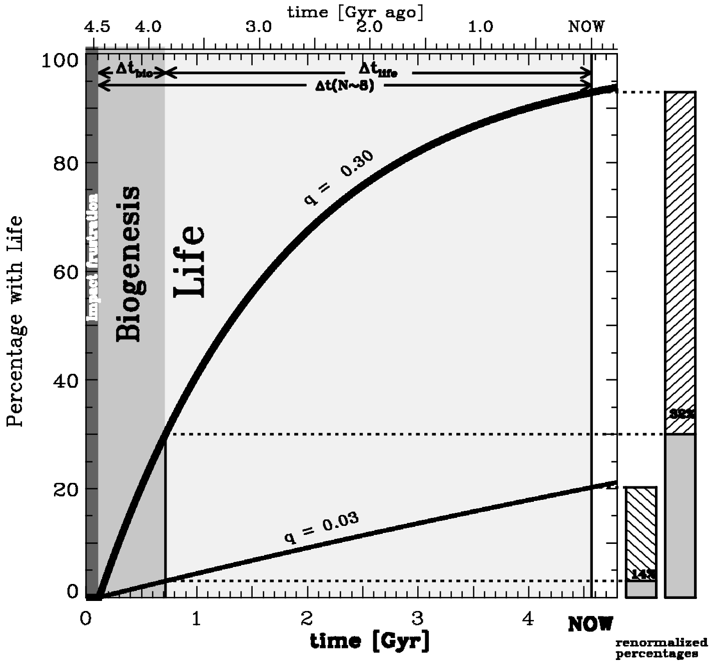

FIG. 2. **We divide the history of the Earth into three epochs: Impact frustration, Biogenesis, and Life.** The black curves show the percentages of terrestial planets with life as a function of time assuming two different probabilities of biogenesis ($q = 0.3, 0.03$), within $\Delta t_{\text{biogenesis}} = 600$ Myr. The percentages in the histograms on the right (14% and 32%) are obtained from comparing the subset of planets that have formed life within $\Delta t_{\text{biogenesis}}$ (middle gray) with the total number of planets that have life (or have had biogenesis) within 4.566 Gyr of their formation (cross-hatched). If q is high (0.30, thick line) a large fraction (32%) of the planets that have evolved life within 4.566 Gyr of formation, have life that evolved rapidly—within $\Delta t_{\text{biogenesis}}$—on their planets. If q is low (0.03, thin line) then a smaller fraction (14%) will have life that evolved rapidly. These different percentages illustrate the principle that a single observation of rapid terrestrial biogenesis is more likely to be the result of high q. This allows us to compute the relative likelihood of q and to constrain q (see Fig. 4).

conditions were probably an effective and recurring autoclave for sterilizing the earliest life forms or more generally frustrating the evolution of life. A steadily decreasing heavy bombardment continued until ~3.8 Gyr ago.

Estimates of the time of the most recent sterilizing impact range between 4.44 and 3.7 Gyr ago (Maher and Stevenson, 1988; Oberbeck and Fogleman, 1989; Sleep *et al.*, 1989; Halliday, 2001). These estimates span the time from the solidification of the Moon's crust to the end of the late heavy bombardment. Thus, life was frustrated for at least the first 0.1 Gyr and possibly as long as the first 0.9 Gyr of the Earth's existence. We take our preferred value as the middle of this range: $\Delta t_{\mathrm{frustrated}} \approx 0.5 \pm 0.4$ Gyr. The range of these estimates reflects the large uncertainties due to small number statistics for the largest impactors and the uncertainty of the energy required to sterilize the Earth completely. We do not know where biogenesis happened or the extent to which it was protected from the effects of impacts. Tidal pools have little protection, hydrothermal vents have some protection, while autotrophic thermophiles in subsurface rock under several kilometers of crust were probably in effective bomb shelters.

The roots of the universal tree of life point to a thermophilic origin (or at least a thermophilic common ancestor) for all life on Earth (Pace, 1991; Stetter, 1996). This suggests a hot biogenesis in hydrothermal vents or possibly subsurface rock and/or selection for thermophilia by periodic temperature pulses from large impacts. *If* we knew that life evolved on the surface of the Earth and was therefore more susceptible to impact sterilizations, life would have been frustrated longer, and our preferred value would be more precise: $\Delta t_{\mathrm{frustrated}} \approx 0.7 \pm 0.2$ Gyr.

If we accept the carbon isotopic evidence for life >3.85 billion years ago (Mojzsis *et al.*, 1996), then life has been on Earth *at least* that long (i.e., Δt_{life} is at least 3.85 Gyr). In addition, because of the Earth's tectonic history, this time is also the earliest time we could reasonably hope to find biological evidence from rocks on Earth—even if life existed earlier. With this selection effect in mind (which we know exists at some level), our preferred value for the time life has existed on Earth is $\Delta t_{\mathrm{life}} \approx 4.0^{+0.4}_{-0.2}$ Gyr.

It is possible that biogenesis occurred several times on Earth. For example, during the period $\Delta t_{\mathrm{frustrated}}$, life could have evolved and been sterilized multiple times. We do not know if this happened. Similar potential sterilizations and biogeneses could have occurred during the period Δt_{life}, but we do not know. For the purposes of this analysis we can ignore this complication. We are only interested in the shortest period within which the observations can constrain biogenesis to have occurred. Thus, $\Delta t_{\mathrm{biogenesis}}$ is our best observational constraint on any epoch of biogenesis, and this happens to be on the period between the most recent sterilizing impact that is older than the oldest evidence we have for life on Earth.

Since our preferred values yield $\Delta t_{\mathrm{frustrated}} + \Delta t_{\mathrm{life}} \approx \Delta t_{\mathrm{Earth}}$, there is little time left for biogenesis to have occurred. This is the basis for the statement that biogenesis occurred rapidly. Substituting our three preferred values, $\Delta t_{\mathrm{Earth}}$, $\Delta t_{\mathrm{frustrated}}$, and Δt_{life}, into Eq. 1 and solving for $\Delta t_{\mathrm{biogenesis}}$ yields:

$$\Delta t_{\mathrm{biogenesis}} = 0.1^{+0.5}_{-0.1}\ \mathrm{Gyr} \tag{2}$$

Thus we take 600 Myr as a crude estimate of the upper limit for the time it took life to appear on Earth. Assuming biogenesis took place on the surface of the Earth, Oberbeck and Fogleman (1989) found this maximum time to be ~25 Myr. Maher and Stevenson (1988) assumed that biogenesis took somewhere between 0.1 and 10 Myr depending on environment, while Sleep *et al.* (2001) found a similar range for evolutionarily significant periods of clement surface conditions. Independently, biologists specializing in the chemistry of the origin of life have estimated that the time required for biogenesis is potentially quite short (Miller, 1982) and could be <8 Myr (Lazcano and Miller, 1994). Thus, several lines of evidence indicate that biogenesis was geologically rapid. If biogenesis occurred in the fissures around a hydrothermal vent or in subsurface rocks, it was well protected, and we can only constrain biogenesis to have taken less than ~600 Myr. If biogenesis occurred on the surface, it probably took less than ~25 Myr. In this analysis we consider 600 and 25 Myr to represent the high and low ends of a plausible range for $\Delta t_{\mathrm{biogenesis}}$ (Figs. 2 and 3, respectively).

These observational constraints on $\Delta t_{\mathrm{frustrated}}$, $\Delta t_{\mathrm{biogenesis}}$, and Δt_{life} are important because they quantify how rapidly biogenesis occurred on Earth and enable us to put limits on the probability that biogenesis occurred on other planets. For example, consider a group of terrestrial planets with approximately the same probability of

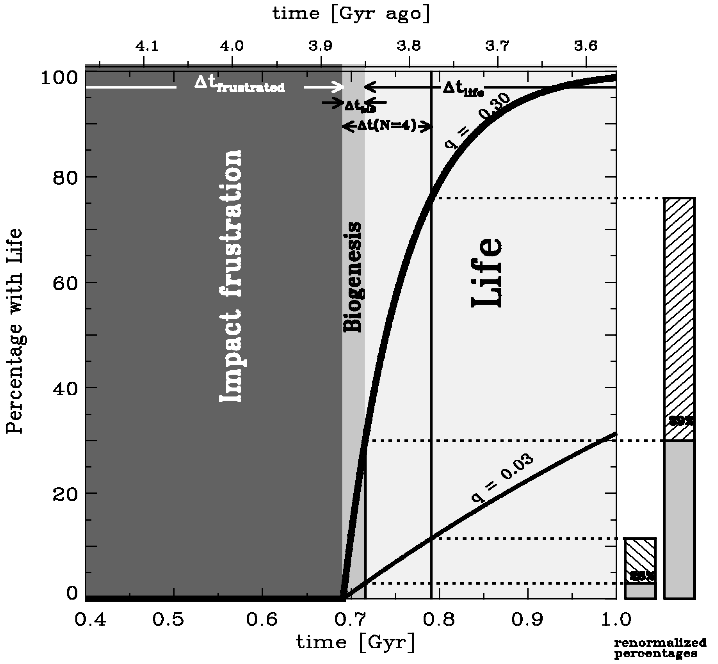

FIG. 3. Zoomed-in version of Fig. 2 with two differences. (1) The window for biogenesis shown here is at the short end of the range permitted by observations: $\Delta t_{\text{biogenesis}} = 25$ Myr. (2) In Fig. 2, by normalizing the histograms to 4.566 Gyr, we have assumed that Earth is a random member of a group of planets at least 4.566 Gyr old upon which biogenesis could have happened *anytime* up to 4.566 after formation, including 1 million years ago. This ignores the nonobservability-of-recent-biogenesis selection effect. Here, to minimize the influence of this selection effect, we only allow biogenesis to occur anytime earlier than 3.77 Gyr ago [i.e., within "$\Delta t(N = 4)$"]. The influence of this normalization time $\Delta t(N)$ is illustrated in Fig. 4.

biogenesis "q" as Earth. Suppose $q = 0.30$. At their formation, none of these planets had life. As time passed, life arose on more and more of them. The thick line in Fig. 2 shows the increasing percentage of these planets with life as time passes. After $\Delta t_{\text{biogenesis}}$, 30% will have life (that is how $q = 0.30$ is defined). After 4.566 Gyr, 93% will have life (7% still will not). Of that 93%, 32% will, like the Earth, have had biogenesis within $\Delta t_{\text{biogenesis}}$. The histogram on the far right of Fig. 2 represents these numbers. Suppose q is only 0.03. Then only 20% will have life after 4.566 Gyr, and only 14% of those will have had biogenesis, like the Earth, within $\Delta t_{\text{biogenesis}}$. Assuming Earth is a random member of the planets with life, the single observation that biogenesis occurred within $\Delta t_{\text{biogenesis}}$ on Earth indicates that larger values of q are more likely than small values. This is the basic idea behind our analysis. It is illustrated in Figs. 1, 4, and 5, which also show quantitative constraints on q under the various assumptions discussed in the next section.

The histograms in Figs. 2 and 3 also show that if q is large, the fraction of gamblers or terrestrial planets who have won after a certain time is large. Suppose the gambler did not know how many gamblers had won by the 12th day, but only that he is a random member of the group that had won. From the fact that he won quickly, he can infer that q is large. This tells the gambler that after a few days a large fraction of lottery ticket buyers are in the winners' group. He is not alone. For the biogenesis lottery, finding large q means that after a few times $\Delta t_{\text{biogenesis}}$, a large fraction of the terrestrial planets with probability of biogenesis similar to the Earth have (or have had) life.

SELECTION EFFECTS

Daily lottery → biogenesis lottery

In the first section we drew parallels between a daily lottery and a biogenesis lottery. Explicitly, these parallels are:

- The first day of the lottery corresponds to the end of $\Delta t_{\text{frustrated}}$, the time when conditions become clement enough for biogenesis.
- All the gamblers had the same (but unknown)

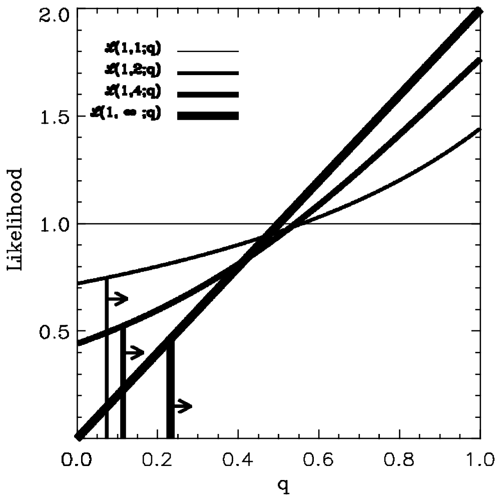

FIG. 4. The effect on the likelihood of varying the somewhat arbitrary number of days that the gamblers have to have won by to be in the group. In contrast to Fig. 1, all the likelihoods of q here are based on the information that a gambler (chosen at random from the group whose members have won within N days) has won on the first day. These likelihoods are plotted and labeled "$\mathscr{L}(1,N;\ q)$" with $N \in \{1,2,4,\infty\}$. In the biogenesis lottery (just as in the daily lottery) N defines the group and is a measure of the duration biogenesis could have taken [$\Delta t(N) = N \times \Delta t_{\text{biogenesis}}$). For example, if $N = 1$ we cannot say anything about q since the likelihood $\mathscr{L}(1,1;\ q)$ is flat (we have conditioned on "rapid" biogenesis). While if $N = \infty$ we can put the strongest constraint on q. If $n = 1$ (when it could have been much larger, $1 \leq n \leq N$), we have more information about q and the likelihood for large q is higher (see Appendix, Eq. A3). The ratio of the probabilities in the histograms in Fig. 3 corresponds to the ratio of the $\mathscr{L}(1,4;\ q)$ likelihoods here: 39%/26% = $\mathscr{L}(1,4;\ q = 0.30)/\mathscr{L}(1,4;\ q = 0.03) = 0.69/0.46 = 1.5$. That is, with $n = 1$ and $N = 4$, values of $q \approx 0.30$ are 50% more likely to be the case than $q \approx 0.03$ [in Fig. 3, $\Delta t(N) = 4 \times 25$ Myr]. Notice that for $2 \leq N \leq \infty$, the 95% lower limit for q varies only between 0.07 and 0.23.

chance of winning the lottery each time they bought a ticket (q is the probability of winning per day). This corresponds to a group of terrestrial planets with approximately the same (but unknown) chance of biogenesis as the Earth (q is the probability of biogenesis within a period of time called here $\Delta t_{\text{biogenesis}}$).

- We selected a gambler at random from those who had won on or before the Nth day. Thus, we conditioned on winning before a certain time. This corresponds to assuming that the Earth is a random member of the group of terrestrial planets that has had biogenesis on or before the end of $\Delta t(N) = N \times \Delta t_{\text{biogenesis}}$. Conditioning on biogenesis before this time corresponds to correcting for the selection effect that biogenesis had to have occurred for us to be here.
- A gambler chosen at random, from the group that has won within N days, found that he had won on the first day ($n = 1$). This corresponds to finding that biogenesis has occurred rapidly on Earth, that is, within $\Delta t_{\text{biogenesis}}$.
- If we set $N = 1$ [see Fig. 4, $\mathscr{L}(1,1;\ q)$], we are conditioning on rapid biogenesis. We are considering a group, all of whose members have had rapid biogenesis. In this case, a random member having rapid biogenesis can tell us nothing about the probability q. To infer something about q we must have $N > 1$.

Nonobservability of recent biogenesis

If our conclusions from the daily lottery are to apply to biogenesis on terrestrial planets we need

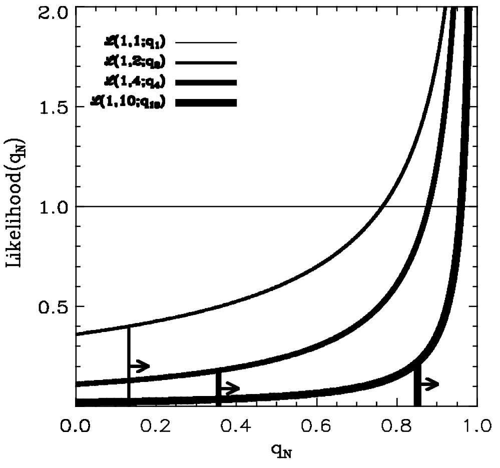

FIG. 5. Likelihood of q_N. In Figs. 1 and 4 the likelihood of q is shown, where q is the probability of winning on any one day. Here we show likelihoods of q_N, the unknown probability of winning on or before the Nth day. This figure shows the effect of varying N. The information used to compute these likelihoods is that a gambler chosen at random from the group whose members have won within N days has won on the first day (see Appendix, Eqs. A6–A8). Translated, this becomes a planet chosen at random, from the group of planets that has had biogenesis within $\Delta t(N) = N \times \Delta t_{\text{biogenesis}}$, has had biogenesis within $\Delta t_{\text{biogenesis}}$. As in Fig. 4, if $N = 1$ we can say nothing meaningful about q_1 ($= q$). However, even if $N = 2$, we can make a stronger statement about q_2 than we could about q: $q_2 > 0.13$ at the 95% confidence level. The 95% lower limit on q_N increases dramatically as N increases, constraining q_N to be close to 1. The ability to extract a useful constraint even if N is low reduces the influence of the selection effects discussed in the text.

to correct for the fact that the evolution of an observer takes some time. How long it takes observing equipment, or complex life or multicellular eukaryotes to evolve, is difficult to say. On Earth it took ~4 Gyr. A limited pace of evolution has prevented us from looking back at our own history and seeing that biogenesis happened last year or even more recently than ~2 Gyr ago (we consider a plausible range for the requisite time elapsed since biogenesis to be between 2 and 10 billion years, or $\Delta t_{\mathrm{evolve}} = 4^{+6}_{-2}$ Gyr).

This selection effect for nonrecent biogenesis is selecting for biogenesis to happen a few billion years before the present regardless of whether it happened rapidly. It is not a selection effect for rapid biogenesis since the longer it took us to evolve to a point when we could measure the age of the Earth, the older the Earth became. Similarly, if biogenesis took 1 Gyr longer than it actually did, we would currently find the age of the Earth to be 5.566 Gyr (= 4.566 + 1) old: "Why is the Earth 4.566 billion years old?" "Because it took that long to find it out." The generalization of this plausible assumption to the ensemble of terrestrial planets is necessary if the likelihoods and constraints in Figs. 1, 4, and 5 are to be applicable to the group of terrestrial planets.

Potential problems

Any effect that makes rapid biogenesis a prerequisite for life would undermine our inferences for q. For example, although it is usually assumed that the heavy bombardment inhibited biogenesis, energetic impacts may have set up large chemical and thermal disequilibria that play some crucial role in biogenesis. We know so little about the details of the chemical evolution that led to life that heat pulses and rapid cooling after large impacts may be part of the preconditions for biogenesis. If true, the time scale of biogenesis would be linked to the time scale of the exponential decay of bombardment, and biogenesis would (if it occurred at all) be necessarily rapid; most extant life in the Universe would have rapid biogenesis, and little could be inferred about the absolute value of q from our sample of 1.

In a panspermia scenario, the rapid appearance of life on Earth is explained not by rapid terrestrial biogenesis as assumed here, but by the ubiquity of the "seeds of life" (e.g., Hoyle and Wickramasinghe, 1999). An analysis in the context of a panspermia scenario would be subject to the same observational constraints as terrestrial biogenesis and would therefore lead to the same inferred probability for the appearance of life on other terrestrial planets.

Another potential problem: Suppose the autocatalytic chemical cycles leading to life are exponentially sensitive to some still unknown peculiarity of the initial conditions on Earth. In this case, to have the same q as the Earth, our group of terrestrial planets may have to be almost indistinguishable rare clones of Earth. That is, conditioning on planets identical to Earth ("same q") would be conditioning on rapid biogenesis ($N = 1$) and would prevent us from inferring much about q from the observations of rapid biogenesis on Earth [see $\mathscr{L}(1,1;\, q)$ in Fig. 4].

This is an example of the more general issue of the status of the rapidity of biogenesis on Earth. Did it have to be rapid? If we assume it could have been otherwise, then we can infer something about q. If it had to be that way, we cannot. The middle ground might be the most plausible option: Biogenesis did not necessarily have to happen as rapidly as it did, but (to be consistent with our existence) it may have had to happen within 1 or 2 billion years of the Earth's formation.

If this is true, we need to look carefully at the influence of varying the somewhat arbitrary and counterfactual duration [$\Delta t(N) = N \times \Delta t_{\mathrm{biogenesis}}$] that biogenesis *could* have taken on Earth. What values of $\Delta t(N)$ are plausible, and how do they affect the results? This is done in Fig. 4, which quantifies the degree of variability one can assume for the duration of biogenesis and still have interesting constraints on q. The lower N is, the less variability is assumed. For example, if $N = 2$ [see $\mathscr{L}(1,2;\, q)$], we are assuming that biogenesis could only have taken as long as $2 \times \Delta t_{\mathrm{biogenesis}}$—enough variability to be able to say something about q but small enough to maintain the nonobservability of recent biogenesis.

Figure 5 allows us to generalize our inferences about q (the probability of biogenesis within $\Delta t_{\mathrm{biogenesis}}$) to inferences about q_N [the probability of biogenesis within an arbitrary time, $\Delta t(N)$]. Specifically, it shows that we only need to be able to assume that biogenesis could have been twice as long as $\Delta t_{\mathrm{biogenesis}}$ to have interesting constraints: $q_2 > 0.13$ at the 95% confidence level [see $\mathscr{L}(1,2;\, q_N)$]. This is the result reported in the abstract.

HOW COMMON IS LIFE?

Relation of our analysis to the Drake Equation

The Drake Equation was devised to address the question of "How many communicative civilizations are in our Galaxy?" (e.g., Sagan, 1973). It has been criticized as "a way of compressing a large amount of ignorance into a small space" (Oliver, cited in Dicke, 1998). Despite its shortcomings, it continues to focus the efforts of the search for extraterrestrial intelligence (SETI) community. An important goal of the SETI community is to turn its subjective probabilities into mathematical probabilities. We have done that here for one of the most important terms.

We are interested in a simpler question: "How common is *life* in the Universe?" Our question is simpler because life is more generic than intelligent or technological life. We modify the Drake Equation to address our question and introduce a parameter F, which is a measure of how common life is in the universe. F is the fraction of stars in our galaxy today orbited by planets that have had independent biogenesis:

$$F = N_l/N_* = f_p\, f_e f_l \qquad (3)$$

where N_l is the number of stars in our Galaxy orbited by planets that have had independent biogenesis, N_* is the number of stars in our Galaxy, f_p is the fraction of stars in our Galaxy with planetary systems, f_e is the fraction of these planetary systems that have a terrestrial planet suitable for life in the same way as the Earth, that is, they have approximately the same probability q as the Earth, and f_l is the fraction of these suitable planets on which biogenesis has occurred.

Many recent observations of the frequency and age dependence of circumstellar disks around young stars in star-forming regions support the widely accepted idea that planet formation is a common by-product of star formation and that the fraction of stars with planetary systems is close to unity, $f_p \approx 1$ (e.g., Habing *et al.*, 1999; McCaughrean *et al.*, 2000; Meyer and Beckwith, 2000).

Equation 3 then becomes:

$$F \approx f_e\, f_l \qquad (4)$$

If $F > 10^{-2}$ then >1% of all stars have (or have had) life, and we conclude that life is "common" in the Universe. If $F < 10^{-11}$, we may be the only life in the Galaxy, and life is "rare."

There has been little agreement on the value of f_l. Hart (1996) wrote, "The value of f_l is extremely speculative," but argued based on the concatenation of low probabilities that it must be extremely small, and thus life elsewhere is improbable. However, Shostak (1998) assumed quite the opposite: "On the basis of the rapidity with which biology blossomed on Earth, we can optimistically speculate that this fraction (f_l) is also one (100%)." Our analysis is a close statistical look at this optimistic speculation.

The relation between the fraction of suitable planets on which biogenesis has occurred ("f_l" in the Drake Equation) and the q analyzed here is:

$$f_l(q,N) = 1 - (1-q)^N \qquad (5)$$

$$f_l(q,\Delta t(N)) = 1 - (1-q)^{\left(\frac{\Delta t(N)}{\Delta t_{\text{biogenesis}}}\right)} \qquad (6)$$

$$f_l(q,t) = 1 - (1-q)^{\left(\frac{t-\Delta t_{\text{frustrated}}}{\Delta t_{\text{biogenesis}}}\right)} \qquad (7)$$

where Eq. 5 refers to the daily lottery (see Appendix, Eq. A5), Eq. 6 is the translation to the biogenesis lottery, and we obtain Eq. 7 by using $t = \Delta t_{\text{frustrated}} + \Delta t(N)$. Equation 7 is plotted in Figs. 2 and 3. Thus, in Eq. 7 we have expressed the "f_l" term of the Drake Equation as a function of time, of the observables $\Delta t_{\text{frustrated}}$ and $\Delta t_{\text{biogenesis}}$, and of the probability q, for which we have derived the relative likelihood. In addition, since $f_l(q,N) = q_N$ (see Eq. A5), Fig. 5 shows the relative likelihood of f_l and establishes quantitative but model-dependent (specifically $N > 1$-dependent) constraints on f_l. For example, for terrestrial planets older than ~1 Gyr (~2 × $\Delta t_{\text{biogenesis}}$), $f_l > 13\%$ at the 95% confidence level.

However, to go from f_l to an answer to the question "How common is life?" (i.e., to go from f_l to F in Eq. 4, we need to know f_e: the fraction of these planetary systems with a terrestrial planet suitable for life in the same way as the Earth. Our understanding of planet formation is consistent with the idea that "terrestrial" planet formation is a common feature of star formation (Wetherill, 1996; Lissauer and Lin, 2000). The mass histogram of detected extrasolar planets peaks at low masses ($dN/dM \propto M^{-1}$) and is also consistent with this idea (Tabachnik and Tremaine, 2001; Zucker and Mazeh, 2001; Lineweaver and Grether, 2002).

Our analysis assumed the existence of a group of terrestrial planets with approximately the same, but unknown ($q \in [0,1]$), probability of biogenesis. It is reasonable to postulate the existence of such a q group since, although we do not know the details of the chemical evolution that led to life, we have some ideas about the factors involved: energy flux, temperatures, the presence of water, planet orbit, residence time in the continuously habitable zone, mass of the planet, atmospheric composition, bombardment rate at the end of planetary accretion and its dependence on the masses of the large planets in the planetary system, metallicity of the prestellar molecular cloud, crust composition, vulcanism, basic chemistry, hydration, pH, and presence of particular clay minerals, amino acids, and other molecular building blocks for life (Lahav, 1999). Since all or many of these physical variables determine the probability of biogenesis, and since the Earth does not seem to be special with respect to any of them (i.e., the Earth probably does not occupy a thinly populated region of this multidimensional parameter space), the assumption that the probability of biogenesis on these planets would be approximately the same as on Earth is plausible. This is equivalent to assuming that q is a "slowly" varying function of environment. If true, f_e and F are not vanishingly small. This can be contrasted with the "exponentially sensitive" case discussed in *Potential problems.*

DISCUSSION

Carter (1983) has pointed out that the time scale for the evolution of intelligence on the Earth (~5 Gyr) is comparable to the main sequence lifetime of the Sun (~10 Gyr). Under the assumption that these two time scales are independent, he argued that this would be unlikely to be observed unless the average time scale for the evolution of intelligence on a terrestrial planet is much longer than the main sequence lifetime of the host star [see Livio (1999) for an objection to the idea that these two time scales are independent]. Carter's argument is strengthened by recent models of the terrestrial biosphere indicating that the gradual increase of solar luminosity will make Earth uninhabitable in a billion years or so—several billion years before the Sun leaves the main sequence (Caldeira and Kasting, 1992; Rampino and Caldeira, 1994).

Our analysis is similar in style to Carter's; however, we are concerned with the appearance of the earliest life forms, not the appearance of intelligent life. Subject to the caveats raised in *Potential problems* and by Livio (1999), the implications of our analysis and Carter's are consistent and complementary: The appearance of life on terrestrial planets may be common, but the appearance of intelligent life may be rare.

SUMMARY

- Our *existence* on Earth does not mean that the probability of biogenesis on a terrestrial planet, q, is large, because if q were infinitesimally small and there were only one life-harboring planet in the Universe, we would, of necessity, find ourselves on that planet. However, such a scenario would imply either that Earth has a unique chemistry or that terrestrial biogenesis has taken a long time to occur. Neither is supported by the evidence we have. Since little can be said about the probability, q, of terrestrial biogenesis from our existence, we assume maximum ignorance: $0 \leq q \leq 1$. We then use the observation of rapid terrestrial biogenesis to constrain q (Fig. 4).
- We convert the constraints on q into constraints on the f_l term of the Drake Equation (the fraction of suitable planets that have life). For terrestrial planets older than ~1 Gyr we find that f_l is most probably close to unity and >13% at the 95% confidence level.
- If terrestrial planets are common and they have approximately the same probability of biogenesis as the Earth, our inference of high q (or high q_N) indicates that a substantial fraction of terrestrial planets have life and thus life is common in the Universe.

However, there are assumptions and selection effects that complicate this result:

- Although we correct the analysis for the fact that biogenesis is a prerequisite for our existence, our result depends on the plausible assumption that *rapid* biogenesis is not such a prerequisite.
- Although we have evidence that the fraction of planets that are "terrestrial" in a broad astronomical sense (rocky planet in the continuously habitable zone) is large, this may be different from the fraction of planets that are "terrestrial" in a more detailed chemical sense.

Although we can make reasonable estimates of what the crusts and atmospheres are made of, without detailed knowledge of the steps of chemical evolution, we cannot be sure that astronomically terrestrial planets have the same q as Earth. That is, the fraction of planets belonging to the Earth's q group is uncertain. Thus, although we have been able to quantify the f_l term of the Drake Equation using rapid biogenesis, our knowledge of the f_e term is still only qualitative and inhibits our ability to draw stronger conclusions about how common life is in the Universe.

APPENDIX: LIKELIHOOD COMPUTATIONS

Let the unknown but constant probability of winning a daily lottery be q. Given the information that a gambler who buys one ticket each day for n days lost on the first $n - 1$ days and won on the nth day, we can compute the likelihood function for q (probability of the data, given the model q):

$$L(n; q) = (1 - q)^{n-1} q \tag{A1}$$

This is equal to the fraction of all gamblers who first experienced $(n-1)$ losses and then won. Given only the information that the gambler won at least once on or before the nth day, the likelihood function for q is:

$$L(\leq n; q) = 1 - (1 - \mathrm{q})^{n} \tag{A2}$$

This is equal to the fraction of all gamblers who have won at least once on or before the nth day. Given the information that a group of gamblers have won at least once on or before the Nth day, and that a gambler chosen at random from this group won at least once on or before the nth day $(n \leq N)$, the likelihood of q is:

$$L(\leq n, \leq N; q) = \frac{L(\leq n; q)}{L(\leq N, q)} = \frac{1 - (1 - q)^{n}}{1 - (1 - q)^{N}} \tag{A3}$$

Out of all the gamblers who have won on or before the Nth day, this is the fraction who have won on or before the nth day. Notice that as $N \rightarrow n$ the likelihood of low q increases and that if $N = n$ the likelihood is the same for all q (see Fig. 4). As $N \rightarrow \infty$ we have $L(\leq n, \leq N; q) \rightarrow L(\leq n; q)$, which yields the tightest constraints on q. A normalized likelihood $\mathscr{L}$ (or probability density) is defined such that $\int_0^1 \mathscr{L}(q)dq = 1$. Thus the renormalization conversion is:

$$\mathscr{L}(x) = \frac{\mathscr{L}(x)}{\int_0^1 L(x)dx} \tag{A4}$$

The normalized likelihoods for Eqs. A1–A3 are plotted in Fig. 1 for the cases $n = 3$ and $N = 12$. The 95% confidence levels cited are Bayesian credible intervals based on a uniform prior for q.

Although q is the probability of winning the lottery in 1 day, we would like to generalize and ask what is the probability q_N of winning the lottery within N days. In the analogous biogenesis lottery, q is the probability of biogenesis on a terrestrial planet with the same unknown probability of biogenesis as the Earth. The time window for biogenesis constrained by observations on Earth, $\Delta t_{\mathrm{biogenesis}}$, corresponds to 1 day for the gambler. However, we would like to compute the likelihood of biogenesis after an arbitrary period of time $\Delta t(N) = N \times \Delta t_{\mathrm{biogenesis}}$. Let q_N be the probability of winning on or before the Nth day:

$$q_N = 1 - (1 - q)^{N} = L(\leq N, q) \tag{A5}$$

(see Eqs. A2 and A3). We would like to know the likelihood of q_N rather than limit ourselves to the likelihood of q $(=q_1)$. Suppose the information is the same that was available to compute $L(\leq n, \leq N; q)$. That information is: A randomly chosen gambler from the group of gamblers who have won after N days, won on or before the nth day. We want $L(\leq n, \leq N; q_N)$. The relationship between the likelihood of q_N and the likelihood of q is

$$\mathscr{L}(\leq n, \leq N; q_{\mathrm{N}})\, dq_{\mathrm{N}} = \mathscr{L}(\leq n, \leq N; q)\, dq \tag{A6}$$

which, with $dq_{\mathrm{N}}/dq = N(1 - q)^{N-1}$ (from Eq. A5), becomes:

$$\mathscr{L}(\leq n, \leq N; q_N)$$

$$= \frac{\mathscr{L}(\leq n, \leq N; q)}{dq_N/dq} \tag{A7}$$

$$= \frac{\mathscr{L}(\leq n, \leq N; q)}{N(1 - q)^{N-1}} \tag{A8}$$

which is plotted in Fig. 5 and has, as expected, relatively larger likelihoods for larger values of q_N.

ACKNOWLEDGMENTS

We acknowledge David Nott and Luis Tenorio for vetting the statistics, Paul Davies, John Leslie, and an anonymous referee for useful comments, and Kathleen Ragan for editing. We thank Don Page for gracefully pointing out an error in the plotting of Eq. A8 in Fig. 5 of a previous version. C.H.L. is supported by an Australian Research Council Fellowship and acknowledges support from the Australian Centre for Astrobiology. T.M.D. acknowledges an Australian Postgraduate Award.

ABBREVIATION

SETI, search for extraterrestrial intelligence.

REFERENCES

Allègre, C.J., Manhès, G., and Göpel, C. (1995) The age of the Earth. *Geochim. Cosmochim. Acta* 59, 1445–1456.

Caldeira, K. and Kasting, J.F. (1992) The life span of the biosphere revisited. *Nature* 360, 721–723.

Canup, R.M. and Asphaug, E. (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412, 708–712.

Carter, B. (1983) The anthropic principle and its implications for biological evolution. *Philos. Trans. R. Soc. Lond.* A310, 347–355.

Charnley, S.B., Rodgers, S.D., Kuan, Y-J., and Huang, H-C. (2002) Biomolecules in the interstellar medium and comets. *Adv. Space Res.* (in press) astro-ph/0104416.

Cronin, J.R. (1989) Origin of organic compounds in carbonaceous chondrites. *Adv. Space Res.* 9(2):59–64.

Dicke, S.J. (1998) *Life on Other Worlds: The 20th-Century Extraterrestrial Debate,* Cambridge University Press, Cambridge, pp. 217–218.

Habing, H.J., Dominik, C., Jourdain de Muizon, M., Kessler, M.F., Laureijs, R.J., Leech, K., Metcalfe, L., Salama, A., Siebenmorgen, R., and Trams, N. (1999) Disappearance of stellar debris disks around main-sequence stars after 400 million years. *Nature* 401, 456–458.

Halliday, A.N. (2000) Terrestrial accretion rates and the origin of the Moon. *Earth Planet. Sci. Lett.* 176, 17–30.

Halliday, A.N. (2001) Earth science: in the beginning. . . . *Nature* 409, 144–145.

Hart, M.H. (1996) Atmospheric evolution, the Drake Equation and DNA: sparse life in an infinite universe. In *Extraterrestrials: Where Are They?*, 2nd edit., edited by B. Zuckerman and M.H. Hart, Cambridge University Press, Cambridge, pp. 215–225.

Hartmann, W.K. and Davis, D.R. (1975) Satellite-sized planetesimals and lunar origin. *Icarus* 24, 504–515.

Hartmann, W.K., Ryder, G., Dones, L., and Grinspoon, D. (2000) The time-dependent intense bombardment of the primordial Earth/Moon system. In *Origin of the Earth and Moon,* edited by R.M. Canup and K. Righter, University of Arizona Press, Tucson, pp. 493–512.

Hoyle, F. and Wickramasinghe, N.C. (1999) *Astronomical Origins of Life: Steps Towards Panspermia,* Kluwer Academic, Boston.

Kasting, J.F., Whitmire, D.P., and Reynolds, R.T. (1993) Habitable zones around main sequence stars. *Icarus* 101, 108–128.

Lahav, N. (1999) *Biogenesis: Theories of Life's Origin,* Oxford University Press, Oxford.

Lazcano, A. and Miller, S.L. (1994) How long did it take for life to begin and evolve to cyanobacteria? *J. Mol. Evol.* 39, 549–554.

Lineweaver, C.H. (2001) An estimate of the age distribution of terrestrial planets in the universe: quantifying metallicity as a selection effect. *Icarus* 151, 307–313.

Lineweaver, C.H. and Grether, D. (2002) The observational case for Jupiter being a typical massive planet. *Astrobiology* 2, 325–334.

Lissauer, J.J. and Lin, D.N.C. (2000) Diversity of planetary systems: formation scenarios and unsolved problems. In *From Extrasolar Planets to Cosmology: The VLT Opening Symposium,* edited by J. Bergeron and A. Renzini, Springer-Verlag, Berlin, p. 377.

Livio, M. (1999) How rare are extraterrestrial civilizations and when did they emerge? *Astrophys. J.* 511, 429–431.

Maher, K.A. and Stevenson, D.J. (1988) Impact frustration of the origin of life. *Nature* 331, 612–614.

McCaughrean, M.J., Stapelfeldt, K.R., and Close, L.M. (2000) High-resolution optical and near-infrared imaging of young circumstellar disks. In *Protostars and Planets IV,* edited by V. Mannings, A.P. Boss, and S.S. Russell, University of Arizona Press, Tucson, pp. 485–507.

Meyer, M.R. and Beckwith, S.V.W. (2000) Structure and evolution of circumstellar disks around young stars: new views from ISO. In *ISO Surveys of a Dusty Universe,* edited by D. Lemke, M. Stickel, and K. Wilke, Springer-Verlag, Heidelberg, pp. 347–355.

Miller, S.L. (1982) Prebiotic synthesis of organic compounds. In *Mineral Deposits and the Evolution of the Biosphere,* edited by W.D. Holland and M. Schidlowski, Springer-Verlag, Berlin, pp. 155–176.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P., and Friend, C.R.L. (1996) Evidence for life on Earth before 3800 million years ago. *Nature* 384, 55–59.

Nealson, K.H. and Conrad, P.G. (1999) Life: past, present and future. *Philos. Trans. R. Soc. Lond. [Biol.]* 354, 1923–1939.

Oberbeck, V.R. and Fogleman, G. (1989) Estimates of the maximum time required to originate life. *Orig. Life Evol. Biosph.* 19, 549–560.

Pace, N.R. (1991) Origin of life—facing up to the physical setting. *Cell* 65, 531–533.

Rampino, M.R. and Caldeira, K. (1994) The Goldilocks problem: climatic evolution and long-term habitability of terrestrial planets. *Annu. Rev. Astron. Astrophys.* 32, 83–114.

Sagan, C. (1973) *Communication with Extraterrestrial Intelligence,* MIT Press, Cambridge, MA.

Shostak, S. (1998) *Sharing the Universe,* Lansdowne, Sydney, p. 180.

Sleep, N.H., Zahnle, K.J., Kasting, J.F., and Morowitz, H.J. (1989) Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature* 342, 139–142.

Sleep, N.H., Zahnle, K., and Neuhoff, P.S. (2001) Initiation of clement surface conditions on the early Earth. *Proc. Natl. Acad. Sci. USA* 98, 3666–3672.

Stetter, K.O. (1996) Hyperthermophiles in the history of life. In *Evolution of the Hydrothermal Ecosystems on Earth (and Mars?), Ciba Foundation Symposium 202,* edited by G.R. Bock and J.A. Goode, John Wiley & Sons, Chichester, UK, pp. 1–10.

Tabachnik, S. and Tremaine, S. (2001) Maximum-likelihood method for estimating the mass and period distributions of extra-solar planets. *Monthly Notices R. Astron. Soc.* 335(1), 151–158.

Trimble, V. (1997) Origin of the biologically important elements. *Orig. Life Evol. Biosph.* 27, 3–21.

Wetherill, G.W. (1996) The formation and habitability of extra-solar planets. *Icarus* 119, 219–238.

Zucker, S. and Mazeh, T. (2001) Derivation of the mass distribution of extrasolar planets with MAXLIMA—a maximum likelihood algorithm. *Astrophys. J.* 562, 1038–1044.

Address reprint requests to:
Dr. Charles H. Lineweaver
School of Physics
University of New South Wales
Sydney, NSW 2052, Australia

E-mail: charley@bat.phys.unsw.edu.au