

Quantifying the information impact of future searches for exoplanetary biosignatures

Amedeo Balbi^{a,1}  and Claudio Grimaldi^{b,c,1} 

^aDipartimento di Fisica, Università di Roma "Tor Vergata," 00133 Roma, Italy; ^bLaboratory of Physics of Complex Matter, Ecole Polytechnique Fédérale de Lausanne, CP-1015 Lausanne, Switzerland; and ^cCentro Studi e Ricerche Enrico Fermi, 00184 Roma, Italy

Edited by Neta A. Bahcall, Princeton University, Princeton, NJ, and approved July 17, 2020 (received for review April 20, 2020)

One of the major goals for astronomy in the next decades is the remote search for biosignatures (i.e., the spectroscopic evidence of biological activity) in exoplanets. Here we adopt a Bayesian statistical framework to discuss the implications of such future searches, both in the case when life is detected and when no definite evidence is found. We show that even a single detection of biosignatures in the vicinity of our stellar system, in a survey of similar size to what will be obtainable in the next 2 decades, would affect significantly our prior belief on the frequency of life in the universe, even starting from a neutral or pessimistic stance. In particular, after such discovery, an initially agnostic observer would be led to conclude that there are more than 10^5 inhabited planets in the galaxy with a probability exceeding 95%. However, this conclusion would be somewhat weakened by the viability of transfer of biological material over interstellar distances, as in panspermia scenarios. Conversely, the lack of significant evidence of biosignatures would have little effect, leaving the assessment of the abundance of life in the galaxy still largely undetermined.

biosignatures | exoplanets | Bayesian analysis | origin of life

Over the past 2 decades, astronomical observations have detected thousands of planets orbiting other stars in our galaxy, allowing us to draw robust statistical conclusions on the populations of such planets (1). Generally speaking, it is now believed that every star in our galaxy should have at least one planet (2) and that many such planets have physical features that may be conducive to the presence of life (3–5).

With the focus of current research rapidly shifting from the detection of exoplanets to their characterization—and, in particular, to the study of their atmospheric composition—we are getting closer to the goal of looking for spectroscopic signatures of biological activity on other worlds (6–9). In the near term, the Transiting Exoplanet Survey Satellite (TESS) (10), CHAracterising ExoPlanet Satellite (CHEOPS) (11), and PLANetary Transits and Oscillations of stars (PLATO) (12) space missions will refine the sample of potentially habitable nearby planets more suitable for follow-up observations. Over the next couple of decades, there will be realistic opportunities for attempting the detection of biosignatures on the most promising targets, both from the ground (e.g., with the European Extremely Large Telescope^{*}) and with dedicated space observatories [such as James Webb Space Telescope (JWST) ref. 13 or Atmospheric Remote-sensing Exoplanet Large-survey (ARIEL) ref. 14]. On a longer time scale, envisioned missions such as Habitable Exoplanet Observatory [Habitable Exoplanet observatory (HabEX[†])], the Large UV/Optical/IR Surveyor[‡], and the Origins Space Telescope[§] might attempt biosignature detection through the direct imaging of habitable rocky exoplanets.

Since technological limitations will initially restrict the search for biosignatures to the immediate vicinity of our stellar system (i.e., within a few tens of light years), a rigorous statistical treatment will be necessary in order to draw conclusions on the possible distribution of inhabited planets in the entire galaxy from a survey of limited spatial extent. This will be true both in the case of

a positive detection of life on one or more exoplanets in a given volume and in the case where no evidence will be found.

Here we suggest an approach to this problem based on the adoption of a Bayesian perspective, showing how existing knowledge or credence on the presence of life beyond Earth will be updated as new evidence will be collected from future missions. A notable previous application of the Bayesian methodology in the context of life emerging in the universe was the attempt to quantify the rate of abiogenesis conditioned on a single datum, i.e., the early appearance of life on Earth, combined with the evidence that it took ≈ 3.8 Gyr for life to evolve intelligence (15). Further developments along this line considered how future evidence would update our previous knowledge on the rate of abiogenesis (16).

Our study tackles the issue of how frequent life is in the universe from a different perspective. We bypass the question of the time scales involved in the abiogenesis, and we rather focus on the present abundance of inhabited planets in the galaxy. In particular, we are interested in assessing the impact of new data (those that could be possibly collected in the next 2 decades) in terms of information gain with respect to existing credence on the probability of life on other planets. We suggest a way to disentangle this unknown probability from others that can be in principle estimated independently, in particular, those pertaining

Significance

The search for life on extrasolar worlds by way of spectroscopic biosignature detection is among the most compelling scientific endeavors of the next decades. This article explores the implications of either discovering or ruling out the presence of detectable biosignatures on planets within a few tens of light years from Earth, a distance within reach of future searches. Using a Bayesian methodology, we show that not detecting biosignatures in such sample volume would bring no added information about the galactic population of life-hosting exoplanets. Conversely, if life arose independently on other planets, even a single detection would imply exobiospheres to be more abundant than pulsars. Putative interstellar transfer of life through the panspermia mechanism may, however, significantly lower this estimate.

Author contributions: A.B. and C.G. designed research, performed research, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: amedeo.balbi@roma2.infn.it or claudio.grimaldi@epfl.ch.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2007560117/-DCSupplemental>.

First published August 17, 2020.

^{*}<https://www.eso.org/sci/facilities/eelt/>.

[†]<http://www.jpl.nasa.gov/habex/>.

[‡]<https://asd.gsfc.nasa.gov/luvoir/>.

[§]<https://asd.gsfc.nasa.gov/firs/>.

to the probability that a specific survey can in fact observe habitable planets. A related relevant question addressed by our study is how hypotheses that assign a lower or higher credence to the presence of life outside Earth—i.e., a pessimistic, neutral, or optimistic attitude toward extraterrestrial life—are weighed and compared in light of new, sparse evidence. Finally, we consider how our results are altered when accounting for the possibility that the distribution of life is correlated over some characteristic distance, such as in panspermia scenarios.

Methods

Main Assumptions. Our statistical model assumes that there are N potentially habitable planets in the Milky Way (i.e., rocky planets orbiting the habitable zone of their host star) and that a survey has looked for spectroscopic biosignatures within a radius R centered around Earth. Statistical estimates based on available data suggest that the percentage of Sun-like (GK-type) and M dwarf stars in our galaxy hosting rocky planets in the habitable zone is about 10 to 20% and 24%, respectively (3–5), resulting in a number of potentially habitable planets of order $N = 10^{10}$. We adopt this estimate as a fiducial value for N in our analysis, without referring to the particular spectral type of the host star (for a recent analysis of this issue, see ref. 17). We further assume that the probability of detecting biosignatures within the survey volume is p , so that the expected number of biosignature detections is

$$\bar{k}(R) = pN\pi(R), \quad [1]$$

where $\pi(R)$ is the probability of a habitable planet being within R , given by

$$\pi(R) = N^{-1} \int d\mathbf{r} \rho(\mathbf{r}) \theta(R - |\mathbf{r} - \mathbf{r}_E|), \quad [2]$$

where \mathbf{r}_E is the position vector of the Earth relative to the galactic center and $\theta(x)$ is the Heaviside step function. The number density function, $\rho(\mathbf{r})$, is defined in such a way that $\rho(\mathbf{r})d\mathbf{r}$ gives the expected number of habitable planets within the volume element $d\mathbf{r}$ about \mathbf{r} , so that $\int d\mathbf{r} \rho(\mathbf{r}) = N$.

The probability p is a shorthand for the various factors that concur to make the presence of detectable biosignatures possible. In our Bayesian analysis we distinguish the factors ascribed to the selection effects of a specific survey from those that are truly inherent to the presence of biosignatures. To this end, we adopt a formalism similar to the one first suggested in ref. 18: this is akin to the Drake equation (19) used in the context of the search for extraterrestrial intelligence but adapted to the search for biosignatures. In our notation, this reduces to writing down the probability p as the product of independent probabilities:

$$p = p_a p_d p_l. \quad [3]$$

The first probability, p_a , pertains to astrophysical factors and observational limitations. Given an exoplanetary survey, only a fraction of systems will be suitable for the search of biosignatures. For example, one may look only for planets in the habitable zone of specific types of stars. The value of p_a can also account for the fact that not all planets in the habitable zone of their star will indeed be habitable. Furthermore, there are other selection effects involved in the specific observational strategy: for example, in a transit survey, there will be strict requirements on the geometrical configuration of the orbital plane, while a direct imaging survey will be limited by the variability of the reflected starlight as the planet orbits the star. In principle, a good estimate of p_a can be obtained from astrophysical and observational considerations. Eventually, a given survey will only sample the quantity $N\pi(R)p_a$. For example, the number of planets that can be scanned for biosignatures following the TESS survey can be estimated to be ≈ 4 , while it would be ≈ 11 for future ground-based imaging (18).

The other probability factors in Eq. 3, p_d and p_l , are not related to a specific survey and pertain exclusively to the likelihood that life-harboring planets in the galaxy display biosignatures. The probability p_d quantifies the fact that in general, detectable biosignatures are not expected to accompany all instances of life on a planet. For example, chemical byproducts of life can significantly alter an exoplanet atmosphere only after some time has passed from the appearance of life. Furthermore, depending on geological and astrophysical factors, life might go extinct after a few hundred million years, as it may have happened on Mars. If we focus on free molecular oxygen as the quintessential biosignature gas, this has been remotely detectable in the Earth atmosphere for ≈ 2 Gyr, roughly half the Earth's age. If we take this as representative of the average, this would point to $p_d \approx 0.5$,

i.e., a large probability. Of course, there is no telling if this is a universal feature of any biosphere, but it nevertheless hints to a significant upper limit on p_d .

Finally, p_l is the probability that life indeed appears on a habitable planet besides Earth. This is, essentially, the probability of abiogenesis and is a truly unknown factor, which makes $\bar{k} = p_d p_l N$, the expected number of planets with biosignatures in the Milky Way, highly indeterminate.

In the present work, we leave unaddressed the possibility of both false negatives (biosignatures that are present but go undetected) and false positives (gases of abiotic origin that are mistakenly interpreted as products of life): however, we note that in principle, both can be incorporated in our formalism through another probability factor, following, for example, the Bayesian framework outlined in refs. 20 and 21 (*SI Appendix, section II*). Our procedure could also be easily specialized for technosignatures, incorporating the appropriate probabilistic factors, along the lines of ref. 22.

In modeling $\pi(R)$ we focus on the thin disk component of the galaxy and adopt an axisymmetric model of the number density of exoplanets:

$$\rho(\mathbf{r}) = N \frac{e^{-r/r_s} e^{-|z|/z_s}}{4\pi r_s^2 z_s}, \quad [4]$$

where r is the radial distance from the galactic center, z is the height from the galactic plane, $r_s = 8.15$ kilo-light years (kly), and $z_s = 0.52$ kly (23). For R smaller than about 1 kly and taking $r_E \approx 27$ kly, the Taylor expansion of $\pi(R)$ for small R yields $\pi(R) \approx (4\pi/3)\rho(\mathbf{r}_E)R^3/N = (R/a)^3$, with $a = 14.2$ kly. Although Eq. 4 assumes that the density profile of habitable exoplanets is proportional to that of stars in the galaxy, other factors such as the metallicity gradient may affect the overall radial dependence of $\rho(\mathbf{r})$ (*SI Appendix, section III*).

Throughout this work, we take an observational radius of $R = 100$ ly which, although corresponding to a galactic fractional volume of only $\pi(R = 100 \text{ ly}) \approx 3.5 \times 10^{-7}$, is an optimistic upper limit of the search range attainable over the next couple of decades. In choosing p_a we consider two limiting situations: 1) $p_a = 1$, which corresponds to an ideal survey that has searched for biosignatures in all of the existing habitable planets within a given distance R from Earth, and 2) $p_a \rightarrow 0$, which corresponds to an exceedingly small number of targeted planets compared to initial sample size (*SI Appendix, section I*).

The probability that a survey searching for biosignature within R finds remotely detectable biosignatures on exactly $k = 0, 1, 2, \dots$ exoplanets follows a binomial distribution:

$$P_k(R) = \binom{N}{k} [\pi(R)p]^k [1 - \pi(R)p]^{N-k}. \quad [5]$$

The average number of exoplanets detectable by the survey is $\bar{k}(R) = Np\pi(R)$, so that by keeping $\bar{k}(R)$ finite, the large N limit of $P_k(R)$ reduces to a Poissonian distribution:

$$P_k(R) = \frac{[\bar{k}(R)]^k}{k!} e^{-\bar{k}(R)}. \quad [6]$$

By rewriting Eq. 1 as

$$\bar{k}(R) = \bar{k} p_a \pi(R), \quad [7]$$

our analysis translates the outcome of a search for biosignatures into an increase in the posterior information on \bar{k} . In practice, we use Bayes theorem to update the prior probability distribution function (PDF) of \bar{k} , after gathering the evidence that exactly k biosignatures are detected in the survey, which is parameterized by $p_a \pi(R)$.

Bayesian Analysis. By isolating the probability factor in p that pertains to astrophysical and observational constraints, p_a , from those referring to the probability of abiogenesis and formation of biotic atmospheres, p_d and p_l , we parameterize the survey by $p_a \pi(R)$ and the expected number of exoplanets in the entire galaxy producing biosignatures by $\bar{k} = p_d p_l N$. Next, we denote \mathcal{E}_k as the event of detecting exactly k biosignatures during the survey, so that using $\bar{k}(R) = \bar{k} p_a \pi(R)$, Eq. 6 gives the likelihood of \mathcal{E}_k being true given \bar{k} :

$$P(\mathcal{E}_k | \bar{k}) = \frac{[\bar{k} p_a \pi(R)]^k}{k!} e^{-\bar{k} p_a \pi(R)}. \quad [8]$$

We aim to find the posterior PDF of \bar{k} resulting from the event \mathcal{E}_k . To this end we consider the prior PDF of \bar{k} , that is, the probability distribution we

ascribe to \bar{k} before gathering the evidence \mathcal{E}_k . In the following, we will refer to a specific functional form of the prior PDF as a model M : $p(\bar{k}|M)$. Following the logic of Bayes' theorem, the posterior PDF is thus obtained from

$$p(\bar{k}|\mathcal{E}_k, M) = \frac{P(\mathcal{E}_k|\bar{k})p(\bar{k}|M)}{P(\mathcal{E}_k|M)}, \quad [9]$$

where

$$P(\mathcal{E}_k|M_i) = \int d\bar{k} P(\mathcal{E}_k|\bar{k})p(\bar{k}|M_i) \quad [10]$$

is the likelihood of \mathcal{E}_k given the model M .

We consider three different models of the prior defined in the interval \bar{k}_{\min} to \bar{k}_{\max} labeled by the subscript $i = 0, 1, 2$:

$$p(\bar{k}|M_i) \propto \bar{k}^{-i}, \quad \text{for } \bar{k}_{\min} \leq \bar{k} \leq \bar{k}_{\max}, \quad [11]$$

where $i = 0$ gives a prior PDF uniform in \bar{k} , which strongly favors large values of \bar{k} (optimistic model M_0), $i = 1$ corresponds to a noninformed prior which is log-uniformly distributed in the interval \bar{k}_{\min} to \bar{k}_{\max} (noninformed model M_1), and $i = 2$ gives a highly informative prior favoring small values of \bar{k} (pessimistic model M_2).

Finally, we consider here only two events resulting from the survey: nondetection, \mathcal{E}_0 , and detection of one biosignature, \mathcal{E}_1 (SI Appendix, section I).

Results and Discussion

Noninformative Prior. We start by assuming no prior knowledge on even the scale of \bar{k} : this is modeled by taking the noninformed log-uniform prior $p(\bar{k}|M_1)$, which gives equal weight to all orders of magnitude of \bar{k} . We take initially $\bar{k}_{\max} = 10^{-2}N = 10^8$, which corresponds to assuming that at most 1 planet out of 100 has detectable biosignatures. In making a choice for \bar{k}_{\min} , one may be tempted to take $\bar{k}_{\min} = 1$ because we know for sure that at least one planet in the Milky Way (the Earth) harbors life. This choice would be justifiable if we were interested in calculating

the posterior PDF of \bar{k} from the evidence gathered within a distance R from a randomly chosen point in the galaxy. However, with the exclusion of the Earth, we ignore whether other planets harbor life (either detectable or not). From our standpoint, therefore, \bar{k}_{\min} can be well below 1. Here we take for the sake of illustration $\bar{k}_{\min} = 10^{-5}$: to give an idea of how small this is, it corresponds to having roughly just one planet with biosignatures in 10^5 hypothetical random realizations of the Milky Way galaxy (we investigate the effect of varying \bar{k}_{\min} below).

Fig. 1A compares the impact of observing or not observing biosignatures within 100 ly. In the case of nondetection, the posterior PDF of \bar{k} differs only marginally from the log-uniform prior (long dashed line) in the range $\bar{k} \lesssim \pi(R)^{-1} \approx 10^6$, even assuming a complete survey ($p_a = 1$). The resulting complementary cumulative distribution function (CCDF) of \bar{k} (Fig. 1D) is somewhat smaller than the corresponding prior CCDF, the main deviation being an upper cutoff for $\bar{k} \gtrsim 10^6$, about 100 times smaller than \bar{k}_{\max} . This limited response to nondetection is explained naturally by the smallness of $\pi(R)$ at $R = 100$ ly, and it becomes even weaker as p_a diminishes, until the prior and posterior probabilities coincide in the entire \bar{k} interval for $p_a \rightarrow 0$. Therefore, even in the hypothesis that future surveys will rule out the existence of detectable biosignatures within 100 ly, the added informative value will nevertheless remain modest, affecting only weakly the initial assertion of a noninformative, log-uniform prior. This conclusion is robust against a lowering of \bar{k}_{\min} and/or \bar{k}_{\max} (SI Appendix, Fig. S8). In particular, reducing \bar{k}_{\max} below $\pi(R)^{-1} \approx 10^6$ is equivalent to assuming that planets with biosignatures are rare enough that finding none within such small survey volume is hardly surprising.

By contrast, the discovery of biosignatures on even a single planet within the entire survey volume ($R = 100$ ly, $p_a = 1$) would

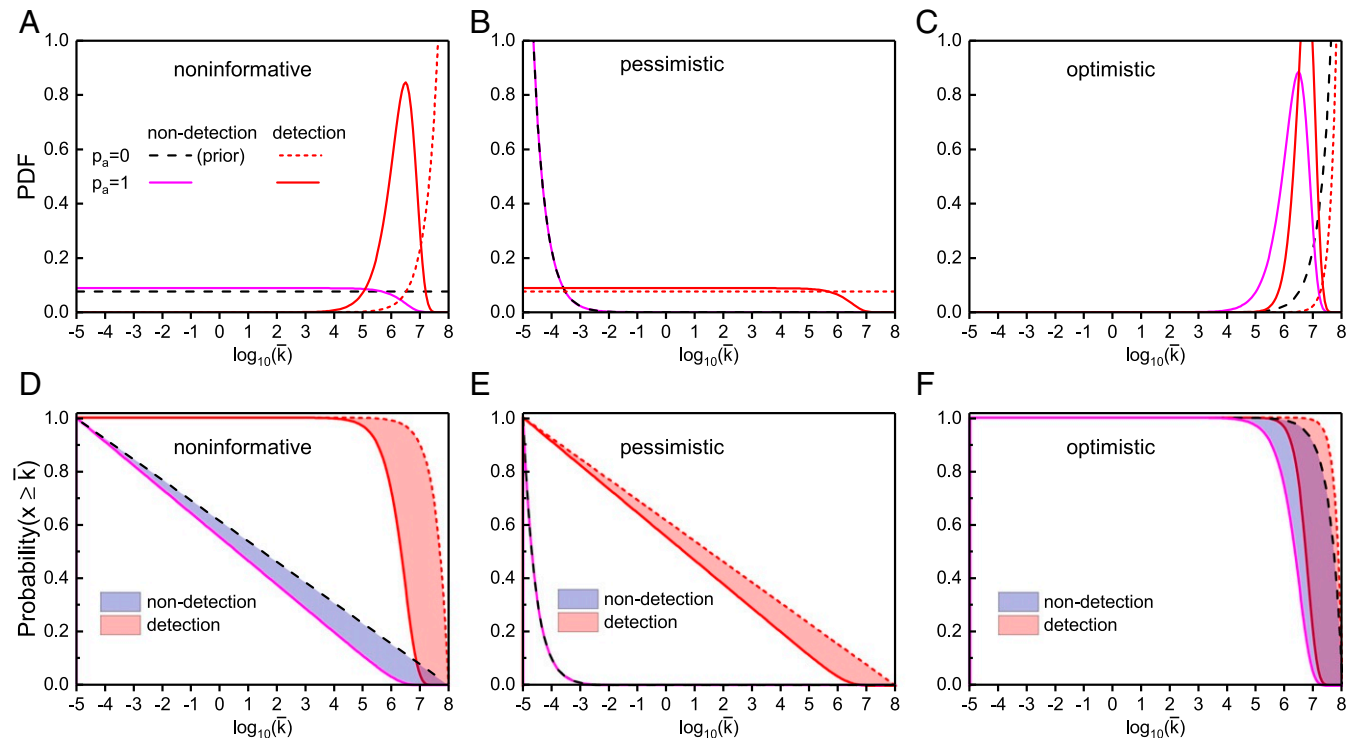


Fig. 1. Results for a survey searching for atmospheric biosignatures within a distance $R = 100$ ly from Earth. Shown are the posterior PDF (A–C) and CCDF (D–F), updated in light of the evidence, starting from a noninformative, pessimistic, and optimistic prior (black dashed curves). The continuous curves refer to the posterior PDF and CCDF for the case $p_a = 1$ (all habitable planets in the survey observed). The limit $p_a = 0$ is shown by red short-dashed curves in the case of detection, while the posteriors resulting from nondetection at $p_a = 0$ coincide with the priors. The shaded areas in the CCDF encompass the limiting cases $p_a = 0$ and $p_a = 1$, giving the range of probabilities that the mean number of life-bearing planets is larger than \bar{k} .

bringing a response markedly different from the prior: we find a posterior PDF strongly peaked around $\bar{k} = 3 \times 10^6$ and a probability exceeding 95% that $\bar{k} > 10^5$. For the sake of comparison, this would imply that exoplanet biosignatures, if distributed homogeneously throughout the galaxy, are far more common than pulsar stars. Even larger values of \bar{k} would be inferred by detecting a biosignature in a sample with few targeted planets, as illustrated by the limiting case $p_a = 0$ in Fig. 1 *A* and *D* (dotted lines) (SI Appendix, section I). We further note that although changing \bar{k}_{\min} does not modify this conclusion, a detection event assuming \bar{k}_{\max} smaller than $\pi(R)^{-1} \approx 10^6$ would bring a response totally independent of the sample fraction p_a , hinting to a larger \bar{k}_{\max} (SI Appendix, Fig. S8).

To provide a more complete analysis of the noninformed case, we have considered also the log-log-uniform prior, which has been designed to reflect total ignorance about the number of conditions conducive to life (24). Although the log-log-uniform PDF slightly favors large values of \bar{k} , the resulting posteriors are in semiquantitative agreement with those resulting from the log-uniform prior of Fig. 1 *A* and *D* (SI Appendix, section IV).

Informative Priors. A log-uniform PDF is probably the best prior reflecting the lack of information on \bar{k} even at the order-of-magnitude level. However, it is also worthwhile to explore how more informative prior distributions are updated once new evidence is gathered. Two interesting limiting cases are those reflecting a pessimistic or optimistic stance on the question of extraterrestrial life. On one hand, it has been argued that abiogenesis may result from complex chains of chemical reactions that have a negligibly low probability of occurring. Furthermore, contingent events which are thought to have favored an enduring biosphere on Earth (like, for example, a moon stabilizing the rotation axis of the planet, plate tectonic, etc.) may be so improbable to further lower the population of biosignature-bearing exoplanets. This view would result in a more pessimistic attitude toward the prior, with small values of \bar{k} being preferred with respect to large ones. We model this case by adopting the uniform in \bar{k}^{-1} prior $p(\bar{k}|M_2) \propto \bar{k}^{-2}$ in the interval \bar{k}_{\min} to \bar{k}_{\max} . Conversely, the astronomically large number of rocky planets in the Milky Way combined with the assumption that the Earth is not special in any way (often termed “principle of mediocrity”) may suggest the optimistic hypothesis that life is very common in the galaxy and the universe, resulting in a prior which weighs large values of \bar{k} more favorably. We capture this view by taking the uniform in \bar{k} prior $p(\bar{k}|M_0)$.

Fig. 1 *B* and *C* show the posterior PDFs and CCDFs resulting from detection or nondetection starting from a pessimistic hypothesis about \bar{k} . While the response to nondetection practically coincides with the prior expectation (an unsurprising result, given that the prior favors small values of \bar{k}), the event of detecting a biosignature increases the cutoff on \bar{k} from $\sim 10^{-3}$ before the detection to at least $\sim 10^6$ after a biosignature is observed within the entire volume sample ($R = 100$ ly, $p_a = 1$). In the optimistic model of Fig. 1 *C* and *F* the prior strongly constrains the posteriors resulting from the events of both detection and nondetection. In particular, the smallness of $\pi(R)$ shifts the CCDF resulting from the nondetection by a factor of only $\sim 10^{-1}$ in \bar{k} (Fig. 1*F*), not justifying thus a substantial revision of the initial optimistic stance.

Model Comparison. By adopting impartial judgement about the probability of M_i being true ($i = 0, 1, 2$), we compute the Bayes factor B_{ij} often used in model selection, giving the plausibility of model M_i compared to M_j in the face of the evidence (i.e., detection or nondetection):

$$B_{ij}(\mathcal{E}_k) = \frac{P(\mathcal{E}_k|M_i)}{P(\mathcal{E}_k|M_j)}. \quad [12]$$

As a reference, $B_{ij} > 10$ is usually considered as strong reason to prefer model M_i over M_j .

Model comparison through the Bayes factor (Fig. 2) shows that if no detection is made, a pessimistic credence with regard to extraterrestrial life would strongly increase its likelihood with respect to an optimistic one, with a Bayes factor above 10, only if p_a is larger than 40%. The increase with respect to a neutral, noninformative stance would be, instead, basically insignificant for all p_a values. This teaches us that unless future surveys will search for biosignatures within a significant fraction of the volume within 100 ly (say, $p_a > 10\%$), detecting none will not support convincingly either hypothesis. On the other hand, if a detection is made, the optimistic scenario would be hugely favored (Bayes factor larger than 10^8) with respect to the pessimistic and would be substantially preferable even with respect to a neutral position when p_a is close to 0. Somewhat counterintuitively, however, finding a single biosignature within a significant fraction of the volume $R = 100$ ly (i.e., p_a larger than 40 to 50%) would not justify entirely the preference for an optimistic credence compared to the noninformative hypothesis. These results put on a quantitative and rigorous statistical basis the common intuitive idea that the discovery of even a single unambiguous

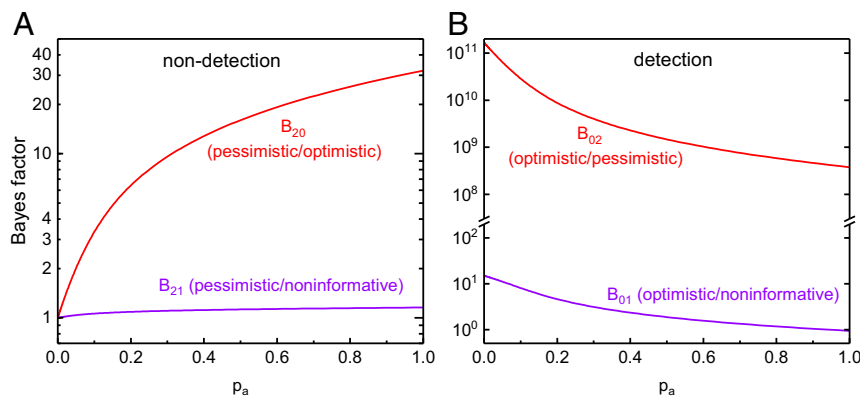


Fig. 2. Bayes factor as a function of p_a , in a survey searching for atmospheric biosignatures within a distance $R = 100$ ly. (A) Bayes factor from the comparison of the pessimistic vs. optimistic model (red) and pessimistic vs. noninformative model (blue), when no biosignature detection is made. (B) Bayes factor from the comparison of the optimistic vs. pessimistic model (red) and optimistic vs. noninformative model (blue), when exactly one biosignature detection is made.

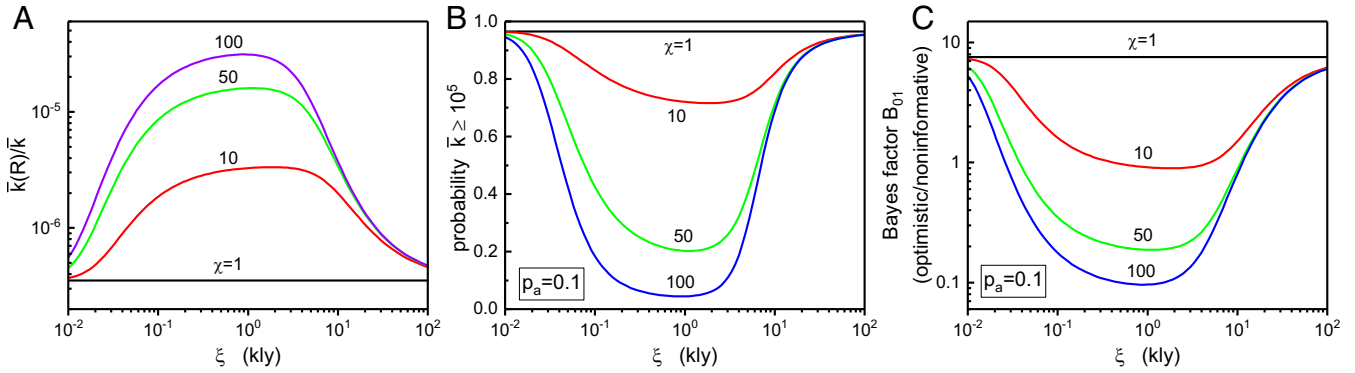


Fig. 3. Effect of assuming correlation between biosignatures due to a panspermia mechanism, after a detection is made in a survey with $R = 100$ ly and $p_a = 0.1$. (A) The fraction of life-harboring planets within a distance R from Earth, $\bar{k}(R)$, over the total number \bar{k} in the entire galaxy, as a function of the correlation length ξ and correlation strength χ . (B) The probability that the total number of life-bearing planets in the galaxy \bar{k} is larger than a reference value 10^5 in the entire galaxy, as a function of the correlation length ξ . (C) Bayes factor from the comparison of the optimistic vs. noninformative model.

biosignature would radically change our attitude toward the frequency of extraterrestrial life.

Impact of Panspermia Scenarios. So far, we have assumed that any given planet has some probability of harboring life independently of whether or not other planets harbor life as well. However, in general, this may not be the case. For example, according to the hypothetical panspermia scenario, life might be transferred among planets, within the same stellar system, in stellar clusters, or over interstellar distances (25–28). If conditions favor the flourishing of a biosphere within a relatively short time scale after the transfer, this would result in an enhanced probability that a planet is inhabited if a nearby planet is inhabited as well (29). In this way, if panspermia can occur, the probability that two planets produce simultaneously biosignatures will depend on their relative distance and on a typical length scale that we denote ξ , defined by the capability of life of surviving transfer and establishing a biosphere.

We took this possibility into account by modeling the statistical correlation of biosignatures and rewriting $\bar{k}(R)$ as follows:

$$\bar{k}(R) = p \int d\mathbf{r} \rho(\mathbf{r}) g(\mathbf{r}, \mathbf{r}_E) \theta(R - |\mathbf{r} - \mathbf{r}_E|), \quad [13]$$

where the pair distribution function $g(\mathbf{r}, \mathbf{r}')$ gives the relative probability of biosignatures being present on \mathbf{r} if biosignatures are present also in \mathbf{r}' .

In principle, different models of correlation could be linked to specific panspermia mechanisms, and various scenarios might even be distinguished observationally from an independent abiogenesis (29). This could be an interesting subject for future studies. However, here we are only interested in how the presence of generic correlations would impact the statistical significance of biosignature detection. We adopt a simple model for the pair distribution function, by assuming that it depends on the relative distance $|\mathbf{r} - \mathbf{r}'|$ in such a way that

$$g(|\mathbf{r} - \mathbf{r}'|) = (\chi - 1)e^{-|\mathbf{r} - \mathbf{r}'|/\xi} + 1, \quad [14]$$

where $\chi \geq 1$ describes the intensity of the panspermia process and ξ describes its spatial extension. The uncorrelated case $g(|\mathbf{r} - \mathbf{r}'|) = 1$ (no panspermia) is obtained by setting $\chi = 1$, while $\chi \gg 1$ yields a strong probability of finding two life-harboring planets within a relative distance $\lesssim \xi$ from each other. Using

$p = p_a p_d p_l$, the average number of exoplanet biosignatures in the entire galaxy is given by $\bar{k} = p_a p_l \int d\mathbf{r} \rho(\mathbf{r}) g(|\mathbf{r} - \mathbf{r}_E|)$, so that Eq. 13 reduces to

$$\bar{k}(R) = \bar{k} p_a \frac{\int d\mathbf{r} \rho(\mathbf{r}) g(|\mathbf{r} - \mathbf{r}_E|) \theta(R - |\mathbf{r} - \mathbf{r}_E|)}{\int d\mathbf{r} \rho(\mathbf{r}) g(|\mathbf{r} - \mathbf{r}_E|)}. \quad [15]$$

The parameter χ is not unbounded, as within any radius R , there cannot be more planets with biosignatures than the total number of planets, $N(R) = N\pi(R)$, contained within R . In other words, $\bar{k}(R) \leq N(R)$ for any R . For $R < 1$ kly, this condition is automatically satisfied by imposing $\chi \leq N/\bar{k}_{\max}$, where \bar{k}_{\max}/N is the maximum fraction of exoplanets harboring biosignatures.

As shown in Fig. 3A, the average number of biosignatures within R , $\bar{k}(R)$, gets enhanced by the panspermia mechanism with respect to the uncorrelated case, with $\bar{k}(R)/\bar{k}$ showing a broad maximum around $\xi = 10^3$ ly (SI Appendix, section IIIB). For much larger values of ξ , panspermia would distribute life homogeneously through the entire galaxy, and $\bar{k}(R)/\bar{k}$ would approach unity.

Fig. 3B shows that if possible correlations in the biosignatures are taken into account, the probability that the number of life-bearing planets in the galaxy is larger than a given value (taken as 10^5 for the sake of illustration) decreases substantially even if life is detected in an incomplete sample (with $p_a = 0.1$) in the volume $R = 100$ ly. This shows as a decrease in the Bayes factor (Fig. 3C) of the optimistic scenario with respect to the noninformative one. Depending on ξ and χ , there may be no gain in knowledge when life is detected elsewhere, and for $\xi \approx 10^3$ ly a complete correlation ($\chi = 100$) would even strongly favor the noninformative hypothesis over the optimistic one. This conclusion suggests that the viability of the panspermia scenario should be assessed independently (for example, through experimental studies of the survivability of organisms in deep space), in order not to weaken the significance of the possible discovery of life beyond Earth.

Data Availability. All data are included in the manuscript and SI Appendix.

ACKNOWLEDGMENTS. A.B. acknowledges support by the Italian Space Agency (DC-VUM-2017-034, Grant 2019-3 U.O. Life in Space) and by Grants FQXi-MGA-1801 and FQXi-MGB-1924 from the Foundational Questions Institute and Fetzer Franklin Fund, a donor-advised fund of Silicon Valley Community Foundation.

1. N. M. Batalha, Exploring exoplanet populations with NASA's Kepler Mission. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12647–12654 (2014).
2. A. Cassan et al., One or more bound planets per Milky Way star from microlensing observations. *Nature* **481**, 167–169 (2012).

3. E. A. Petigura, A. W. Howard, G. W. Marcy, Prevalence of Earth-size planets orbiting sun-like stars. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19273–19278 (2013).
4. J. K. Zink, B. M. S. Hansen, Accounting for multiplicity in calculating eta Earth. *Mon. Not. R. Astron. Soc.* **487**, 246–252 (2019).

5. C. D. Dressing, D. Charbonneau, The occurrence of potentially habitable planets orbiting M dwarfs estimated from the full Kepler dataset and an empirical measurement of the detection sensitivity. *Astrophys. J.* **807**, 45 (2015).
6. S. Seager, The future of spectroscopic life detection on exoplanets. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12634–12640 (2014).
7. J. L. Grenfell, A review of exoplanetary biosignatures. *Phys. Rep.* **713**, 1–17 (2017).
8. Y. Fujii, D. Angerhausen, R. Deitrick, Exoplanet biosignatures: Observational prospects. *Astrobiology* **18**, 739–778 (2018).
9. N. Madhusudhan, Exoplanetary atmospheres: Key insights, challenges, and prospects. *Annu. Rev. Astron. Astrophys.* **57**, 617–663 (2019).
10. G. R. Ricker *et al.*, Transiting exoplanet survey satellite. *J. Astronomical Telesc. Instrum. Syst.* **1**, 014003 (2014).
11. C. Broeg, W. Benz, A. Fortier, “The CHEOPS mission—Rationale and status” in *42nd COSPAR Scientific Assembly* (2018), vol. 42, p. E4.1-5-18.
12. H. Rauer *et al.*, The PLATO 2.0 mission. *Exp. Astron.* **38**, 249D330 (2014).
13. J. P. Gardner *et al.*, The James Webb Space Telescope. *Space Sci. Rev.* **123**, 485–606 (2006).
14. G. Tinetti *et al.*, A chemical survey of exoplanets with Ariel. *Exp. Astron.* **46**, 135–209 (2018).
15. D. S. Spiegel, E. L. Turner, Bayesian analysis of the astrobiological implications of life's early emergence on earth. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 395–400 (2012).
16. J. Chen, D. Kipping, On the rate of abiogenesis from a Bayesian informatics perspective. *Astrobiology* **18**, 1574–1584 (2018).
17. J. Haqq-Misra, R. K. Kopparapu, E. T. Wolf, Why do we find ourselves around a yellow star instead of a red star?. *Int. J. Astrobiol.* **17**, 77–86 (2018).
18. S. Seager, The search for habitable planets with biosignature gases framed by a “Biosignature Drake Equation.” *Int. J. Astrobiol.* **17**, 294–302 (2018).
19. F. D. Drake, “The radio search for intelligent extraterrestrial life” in *Current Aspects of Exobiology*, G. Mamikunian, M. H. Briggs, Eds. (Elsevier, 1965), pp. 323–345.
20. S. I. Walker *et al.*, Exoplanet biosignatures: Future directions. *Astrobiology* **18**, 779–824 (2018).
21. D. C. Catling *et al.*, Exoplanet biosignatures: A framework for their assessment. *Astrobiology* **18**, 709–738 (2018).
22. M. Lingam, A. Loeb, Relative likelihood of success in the search for primitive versus intelligent extraterrestrial life. *Astrobiology* **19**, 28–39 (2019).
23. A. Misiriotis, E. Xilouris, J. Papamastorakis, P. Boumis, C. Goudis, The distribution of the ISM in the Milky Way—A three-dimensional large-scale model. *Astron. Astrophys.* **459**, 113–123 (2006).
24. B. C. Lacki, The log log prior for the frequency of extraterrestrial intelligence, arXiv:1609.05931 (19 September 2016).
25. H. J. Melosh, Exchange of meteorites (and life?) between stellar systems. *Astrobiology* **3**, 207–215 (2003).
26. F. C. Adams, D. N. Spergel, Lithopanspermia in star-forming clusters. *Astrobiology* **5**, 497–514 (2005).
27. I. Ginsburg, M. Lingam, A. Loeb, Galactic panspermia. *Astrophys. J.* **868**, L12 (2018).
28. H. W. Lin, A. Loeb, Statistical signatures of panspermia in exoplanet surveys. *Astrophys. J. Lett.* **810**, L3 (2015).
29. M. Lingam, Analytical approaches to modelling panspermia-beyond the mean-field paradigm. *Mon. Not. R. Astron. Soc.* **455**, 2792–2803 (2016).