*A Bayesian Analysis of the Astrobiological Implications of the Rapid Emergence of Life on the Early Earth**

Ed Turner

Princeton University, Department of Astrophysical Sciences

and

University of Tokyo, IPMU

++++++++++++++++++++++++++++

Harvard University, Statistics Department – November 15, 2011

*Collaboration with **David Spiegel** (Princeton University Observatory)
(+ thanks to former Princeton undergraduate Carl Boettiger'07)

Preface

- Field: Astrobiology
- Content: Mostly statistics with small amounts of biology & geology but *almost no astronomy*
- Motivation 1: Clarify the implications of the early appearance of life on Earth for *biology elsewhere*
- Motivation 2: A worked example of a Bayesian approach to a hand-waving, intuitive, ambiguous, judgment-call, *small-number-statistics* problem with some parallels to *anthropic arguments*

Fundamental Questions of Astrobiology

- Does extraterrestrial life exist?
- What is its nature?
- How common is it?

The overwhelmingly most practical and promising approaches to these questions are empirical, searches for life on bodies in the Solar System and beyond. But that is not our topic today.

A Possible-in-Principle Calculation

Given

- $P_{abiogenesis}$*(physical conditions)* per unit volume and per unit time, and
- prevalence of such suitable abiotic enviroments in the Universe

a straightforward statistical estimate could be made.

But, alas, we are quite ignorant of both and have no immediate prospects of remedying either situation.

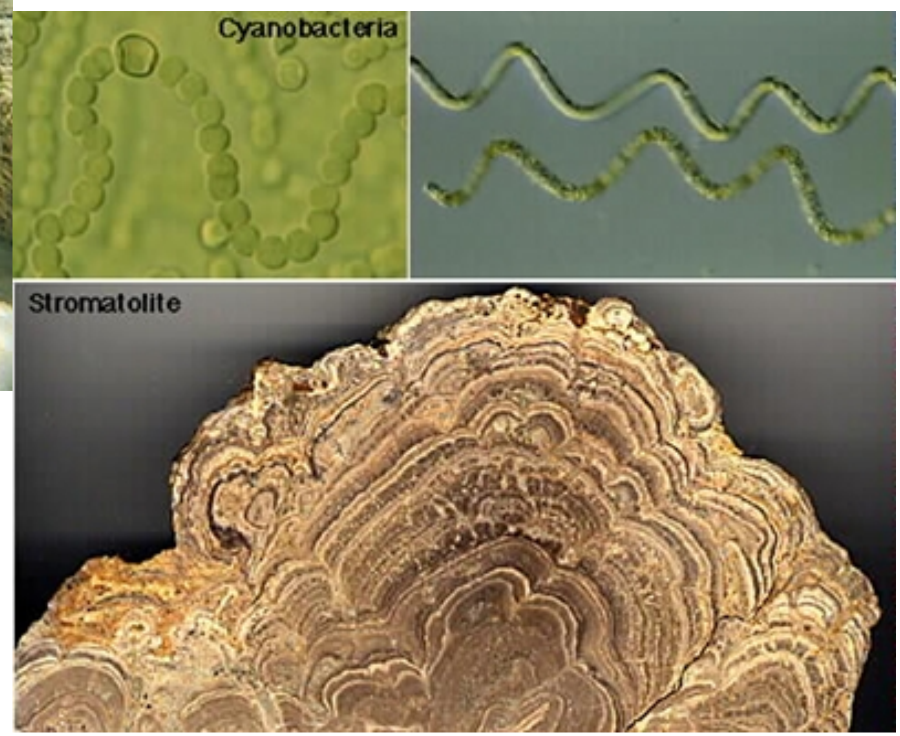
A Potential Finesse

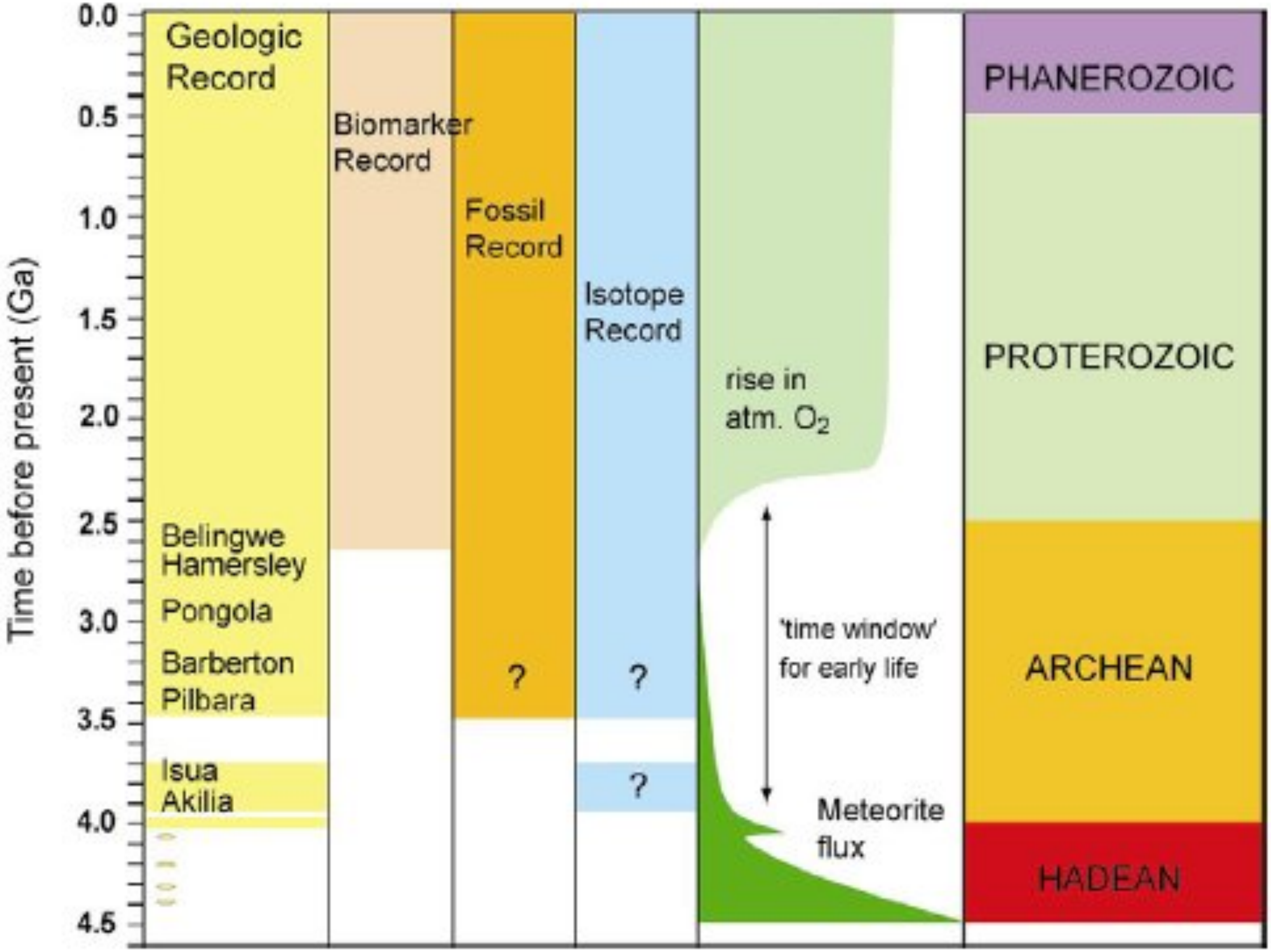
- Exoplanet studies strongly suggest, but have not yet proven, that planets resembling the Earth in a very general/crude way (mass, primary star, orbit, composition etc) are reasonably common.
- Life arose ***very quickly*** on the early Earth.
- **This suggests that $P_{abiogenesis}$*(early-Earth-like conditions)* per unit planet per unit age of the Earth is *not extremely small.***
- Thus, simple/primitive extraterrestrial life (at least) is not too rare. ← roughly the current "consensus" view

How quickly is *very quickly*?

- Earth formed at 4.54 Ga (Ga = 10^9 yrs ago, like redshift)
- Water was reasonably abundant by ~4.3 Ga
- Likely sterilization by the LHB and subsequent global volcanic resurfacing around 3.85-4.0 Ga
- Highly controversial isotopic indicators (^{13}C depletion) of metabolic activity in the oldest known surviving rocks, 3.7-3.82+ Ga
- Wide-spread "probable fossils" of microbes & macroscopic microbial biofilms (stromatolites) in sediments formed ~3.5 Ga
- Definite, highly evolved macro-fossils ~3.2 Ga

Stromatolites





(Van Zuilen 2006)

Abiogenesis almost certainly occurred on Earth significantly earlier than the time of the most ancient life detected to date.

Thus, "*very quickly*" must mean within a few hundred million years but it could have been much faster.

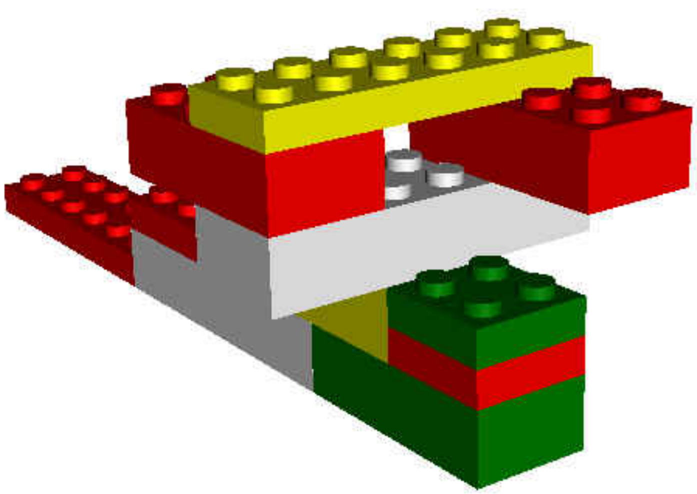
How small is ***extremely small***? (1)

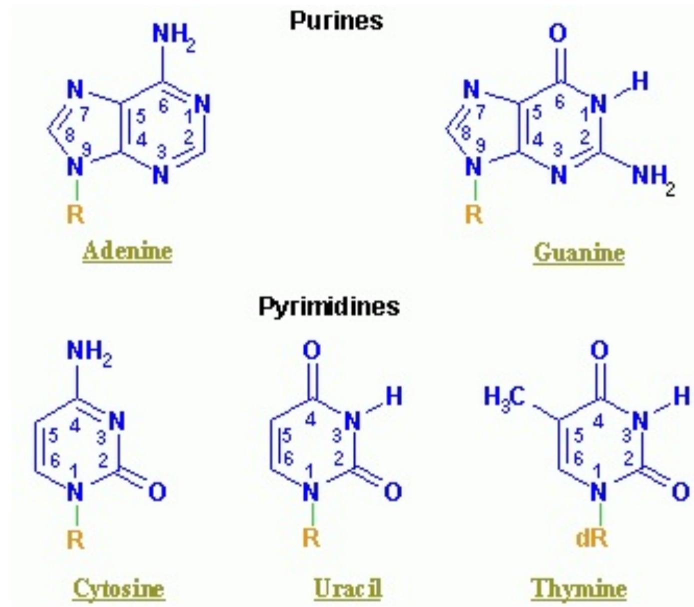
- The basic molecular chemistry of all terrestrial life is essentially identical and *very complex.*
- Two classes of macro-molecules, proteins and nucleic acids, play central roles currently, but it is imagined that RNA alone might have sufficed for an earlier form of life (the "RNA world").
- The building blocks of both proteins (amino acids) and nucleic acids (nucleotides) are a set of small molecules *which form bonds in an arbitrary order* to create the long polymers which are these macro-molecules.

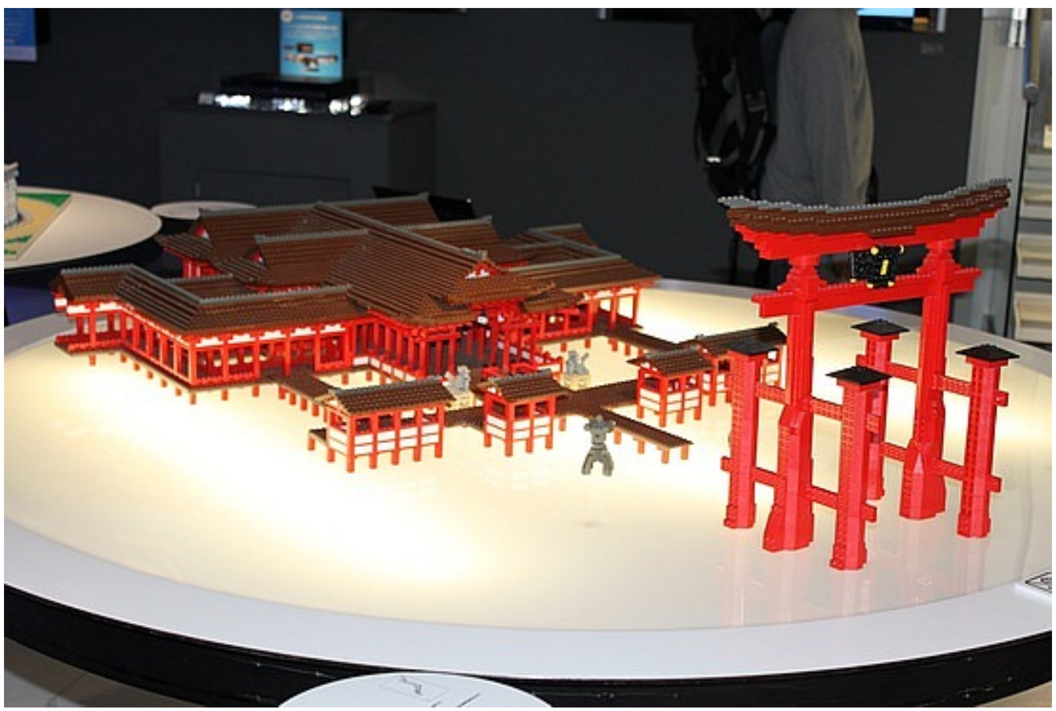
How small is *extremely small*? (2)

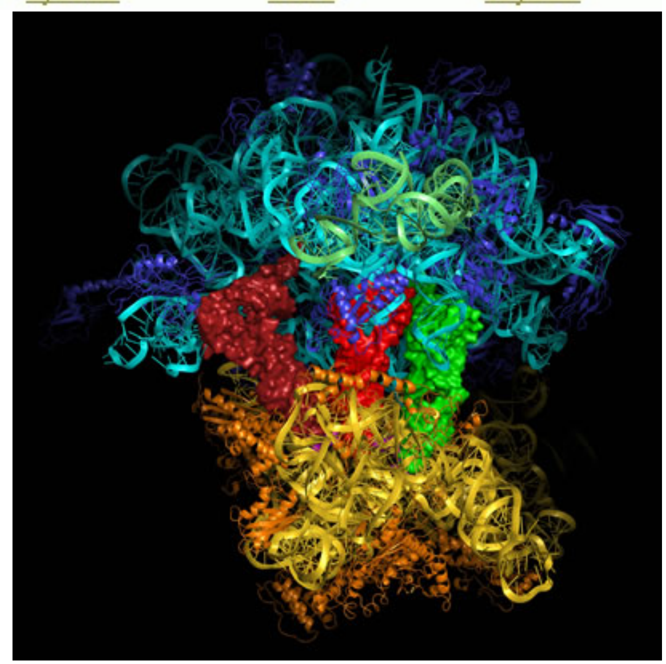
- Both amino acids and nucleotides are produced in reasonable abundance by physical (abiotic) chemical reactions that plausibly (empirically in some cases) commonly occur in nature (GOOD!).
- However, the polymers in question are *huge*, typically 100s to 100s of millions of “building blocks” long, with even minimally functional ones requiring ~100 nucleotides or amino acids
- The probability of any specific one arising randomly is then *factorially small, i.e, ≈10^{-100s} or less (BAD!).*
- Thus, “*extremely small*” ➔ no extraterrestrial life

Abiogenesis: Stochastic Lego Construction?









Ward & Brownlee (2000)
RARE EARTH

"the time from soup to bugs may have been far less than 10 million years. Making life may be a rapid operation – a key observation supporting our contention that life may be very common in the Universe."

Irrational Exuberance

"Personally, given the ubiquity and propensity of life to flourish wherever it can, I would say that, my own personal feeling is that the chances of life on this planet are 100%. I have almost no doubt about it."

- Steven Vogt speaking to L.-J. Zgorski (NSF) in regard to Gl 581g, a reported exoplanet with $Msin(i) \approx 3M_{\odot}$ orbiting within its primary's HZ

Sensible (but not exactly right)

“It is too glib to claim that if the origin of life took place on Earth immediately after the end of the heavy bombardment, then the origin of life must be ‘easy,’ so that the prospects for life elsewhere are increased”

- Chyba & Hand (2005) **ARAA, 43** *Astrobiology: The Study of the Living Universe*

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 ~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.

N.B. – A formal *frequentist* analysis with a misleading result

Start from the Bayesian's *F=ma*

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[B|A] \times \mathrm{P}_{\mathrm{prior}}[A]}{\mathrm{P}[B]} \,.$$

A = the model

B = the data

P[A|B] = probability of the model given the data = the posterior
(what we want!)

P[B|A] = probability of the data given the model = the likelihood
(what we can obtain from a "forward" calculation)

P_{prior}[A] = *a priori* probability the model is true/correct = the prior
(often requires "arbitrary" judgment calls, source of controversy)

P[B] = probability of the data
(typically unknown, but only needed for normalization)

A Uniform Rate (Poisson) Model

λ = rate of abiogenesis per Gyr per Earth-like planet

t = age of the Earth (like expansion factor)

$\tau_{sterile}$ = earliest time at which life could develop

τ_{max} = latest time at which life could develop

n = number of independent abiogenesis events

$$\mathrm{P}[\lambda, n, t] = \mathrm{P}_{\mathrm{Poisson}}[\lambda, n, t] = e^{-\lambda(t-\tau_{\mathrm{sterile}})} \frac{\{\lambda(t - \tau_{\mathrm{sterile}})\}^n}{n!}$$

Two "Anthropic" Constraints

- Life <u>must</u> have arisen on Earth ≥ one times or we would not be here to carry out the calculation

$$P_{\mathrm{life}} = 1 - P_{\mathrm{Poisson}}[\lambda, 0, t] = 1 - e^{-\lambda(t-\tau_{\mathrm{sterile}})}$$

- Life <u>must</u> have arisen early enough on Earth to allow us to evolve by the present

$$t_{\mathrm{emerge}} < t_0 - \tau_{\mathrm{evolve}}$$

t_{emerge} = Earth's age at abiogenesis, t_0 = its present age & τ_{evolve} = time for evolution of astrobiologists

The Likelihood Term

$$\mathrm{P}[B|A] = \frac{1 - \exp[-\lambda(t_{\mathrm{emerge}} - \tau_{\mathrm{sterile}})]}{1 - \exp[-\lambda(\tau_{\mathrm{req}} - \tau_{\mathrm{sterile}})]}$$

where $\tau_{\mathrm{req}} \equiv \min[t_0 - \tau_{\mathrm{evolve}}, \tau_{\mathrm{max}}]$

In the limit of extremely small λ

$$\mathrm{P}[B|A] \approx \frac{t_{\mathrm{emerge}} - \tau_{\mathrm{sterile}}}{\tau_{\mathrm{req}} - \tau_{\mathrm{sterile}}}$$

In the limit of extremely large λ

P[B|A] ≈ 1

The Bayes Factor/Ratio/Evidence for Model Selection

$$\mathcal{R} \equiv \frac{\mathrm{P}[\mathrm{data}|\mathrm{large}\ \lambda]}{\mathrm{P}[\mathrm{data}|\mathrm{small}\ \lambda]} \approx \frac{\Delta t_2}{\Delta t_1}\,.$$

where

$$\Delta t_1 \equiv t_{\mathrm{emerge}} - \tau_{\mathrm{sterile}} \text{ and } \Delta t_2 \equiv \tau_{\mathrm{req}} - \tau_{\mathrm{sterile}}$$

Rule of thumb: $\mathcal{R} < 10$ ➔ barely worth mentioning

The Prior Factor – Problematic!

$$\mathrm{P_{prior}}[A] = f_\lambda[\lambda] f_{\tau_\mathrm{s}}[\tau_\mathrm{sterile}] f_{\tau_\mathrm{m}}[\tau_\mathrm{max}] f_{\tau_\mathrm{e}}[\tau_\mathrm{evolve}]$$

Uniform: $$f_{\lambda,u}[\lambda] = \frac{1}{\lambda_\mathrm{max}}$$ ← L&D2002 equivalent

Informative=BAD

Inverse Uniform: $$f_{\lambda,iu}[\lambda] = \frac{1}{\lambda^2} \times \frac{1}{\lambda_\mathrm{min}{}^{-1} - \lambda_\mathrm{max}{}^{-1}}$$

Informative=BAD

Log Uniform: $$f_{\lambda,l}[\lambda] = \frac{1}{\lambda} \times \frac{\log_{10} e}{\log_{10}\lambda_\mathrm{max} - \log_{10}\lambda_\mathrm{min}}$$

Uniformative=GOOD

but improper=not so good

λ_{max} = 1000 Gyr^{-1} λ_min: 10^{-22} Gyr^{-1}, 10^{-11} Gyr^{-1}, and 10^{-3} Gyr^{-1}

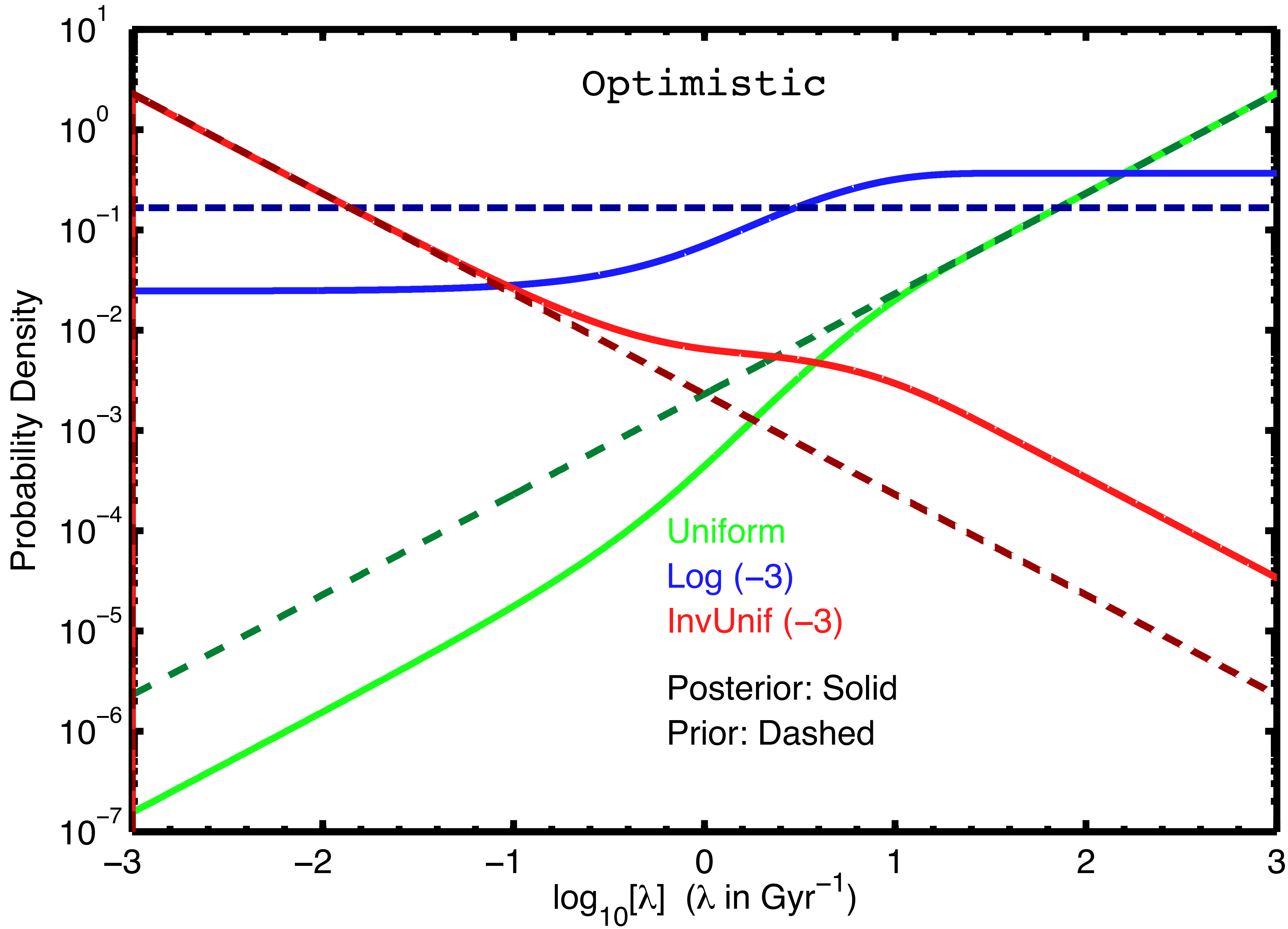
Use δ-function estimates for the f_τ terms

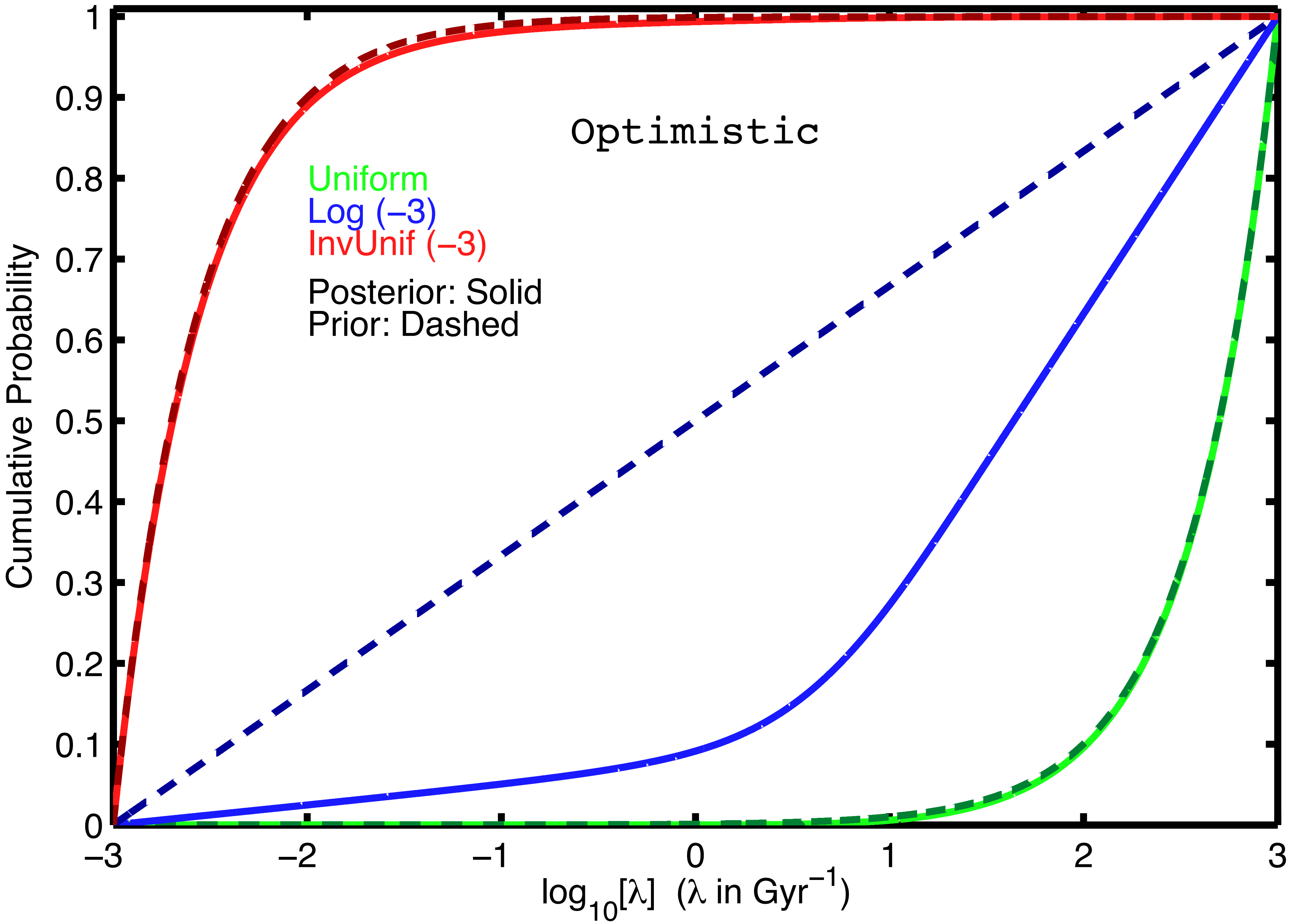
Table 1: Models of $t_0 = 4.5$ Gyr-Old Planets

Model	$\tau_{\rm sterile}$ (Gyr)	$t_{\rm emerge}$ (Gyr)	$\tau_{\rm max}$ (Gyr)	$\tau_{\rm evolve}$ (Gyr)	$\tau_{\rm req}$ (Gyr)	Δt_1 (Gyr)	Δt_2 (Gyr)	$\mathcal{R}$
Hypothetical	0.5	0.51	10	1	3.5	0.01	3.00	300
$\texttt{Conservative}_1$	0.5	1.3	1.4	2	1.4	0.80	0.90	1.1
$\texttt{Conservative}_2$	0.5	1.3	10	3.1	1.4	0.80	0.90	1.1
Optimistic	0.5	0.7	10	1	2.5	0.20	3.00	15

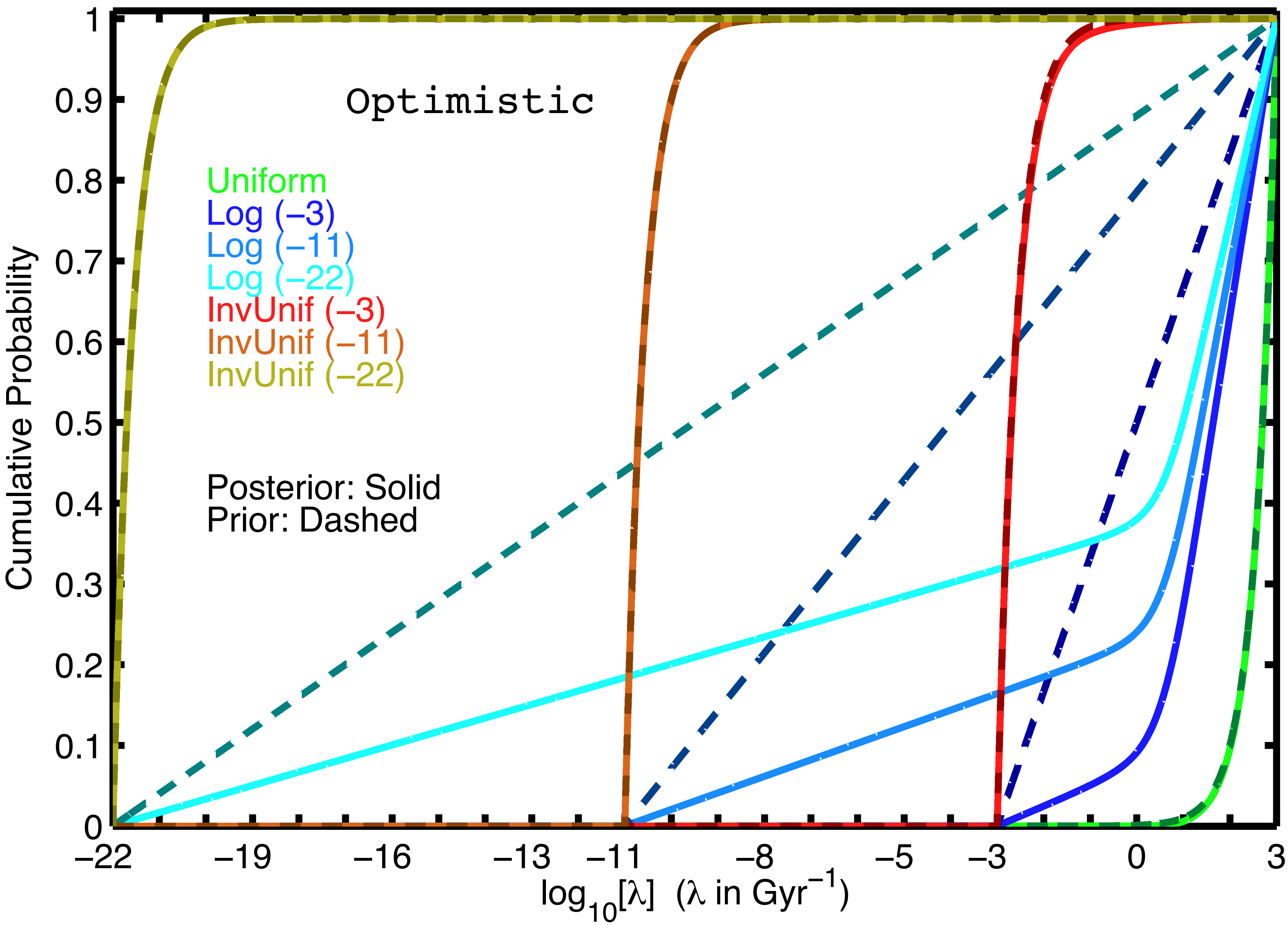
Hypothetical = *extremely* quick terrestrial abiogenesis (speculative)
Conservative = slow terrestrial abiogenesis consistent with the data
Optimistic = quick terrestrial abiogenesis suggested by the data

Turn the Bayes crank ➔ posterior distributions

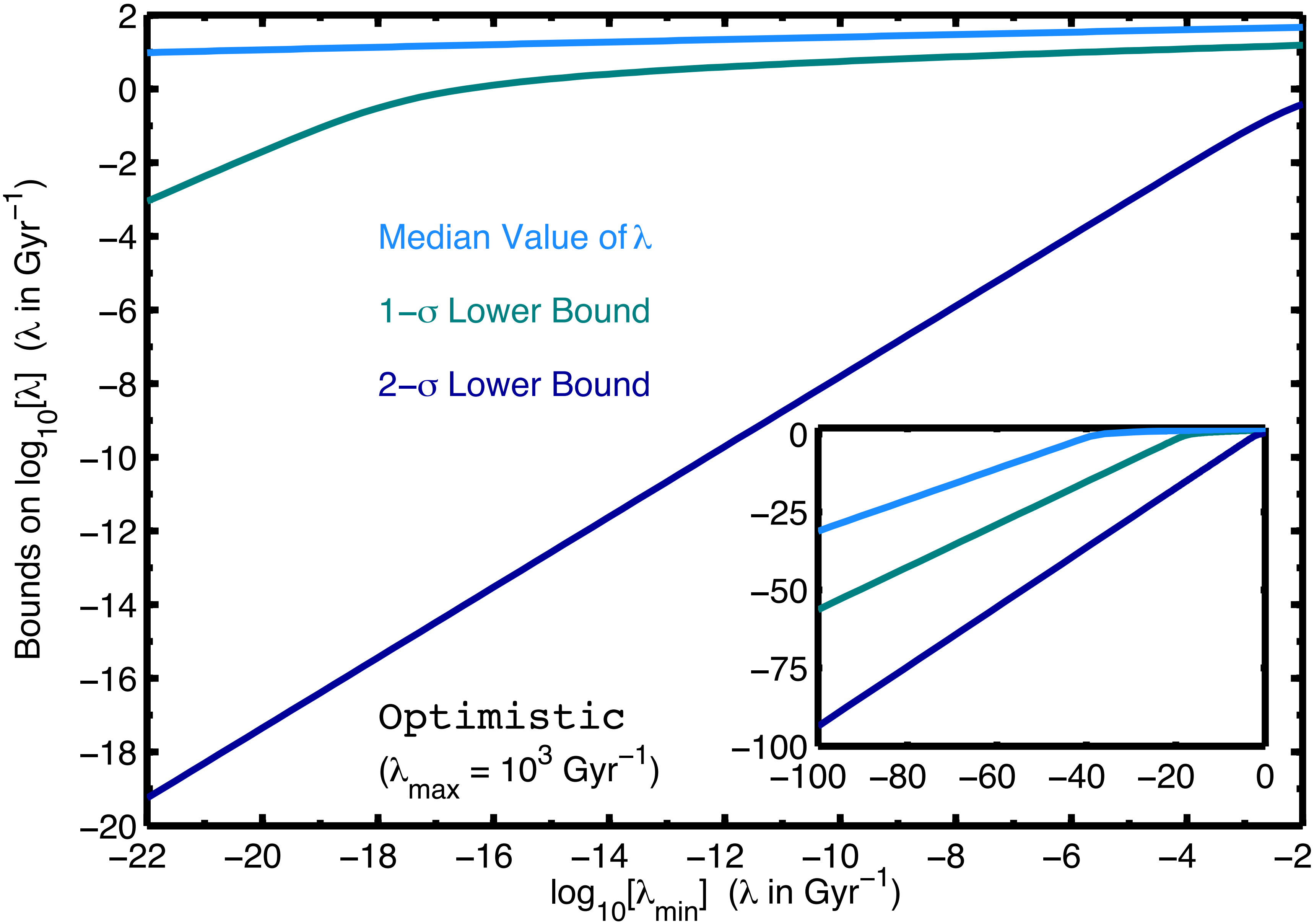

Optimistic
Uniform
Log (–3)
InvUnif (–3)
Posterior: Solid
Prior: Dashed
Probability Density
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
–3
–2
–1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})



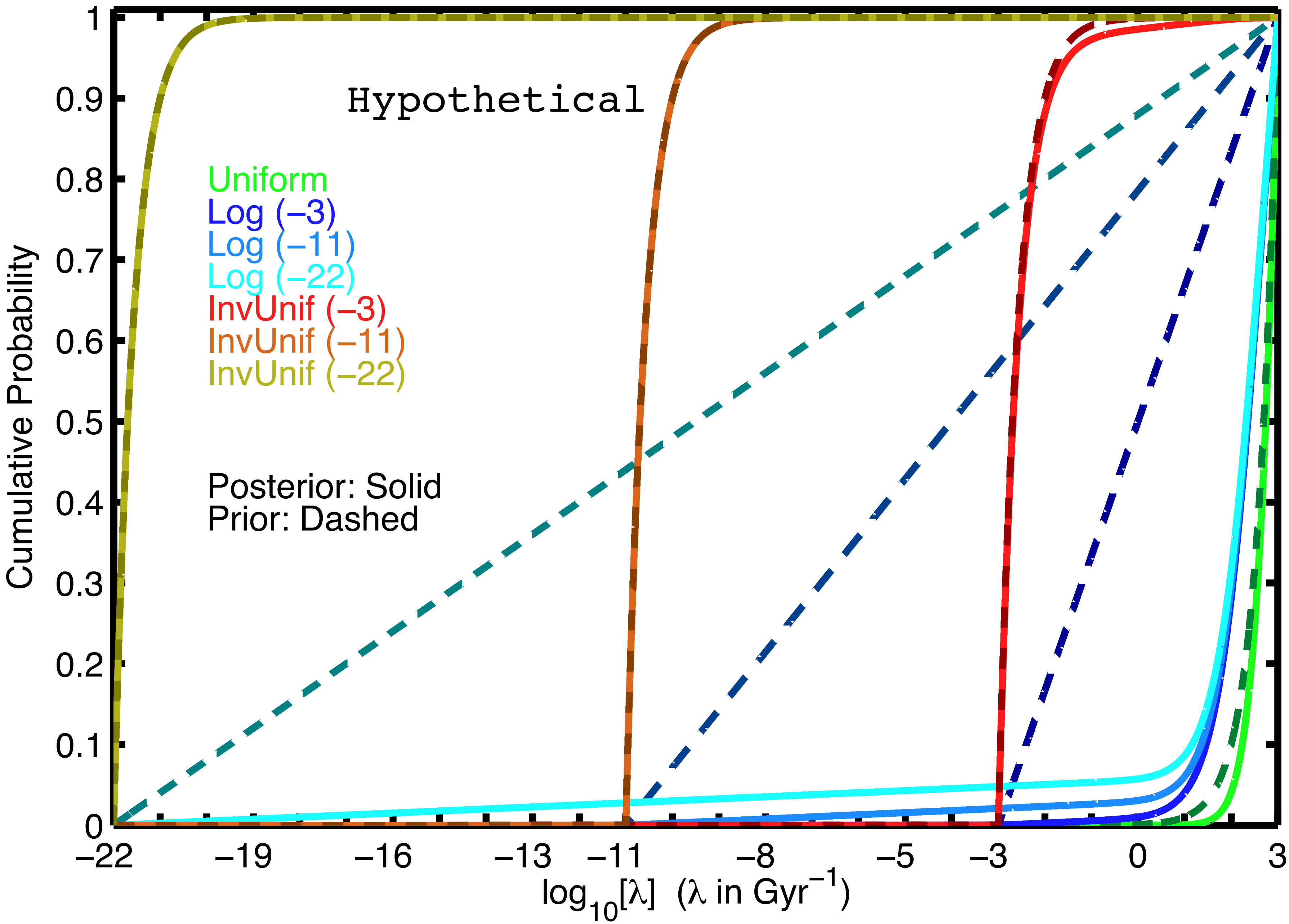
Optimistic
Uniform
Log (–3)
InvUnif (–3)
Posterior: Solid
Prior: Dashed
Cumulative Probability
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
–3
–2
–1
0
1
2
3



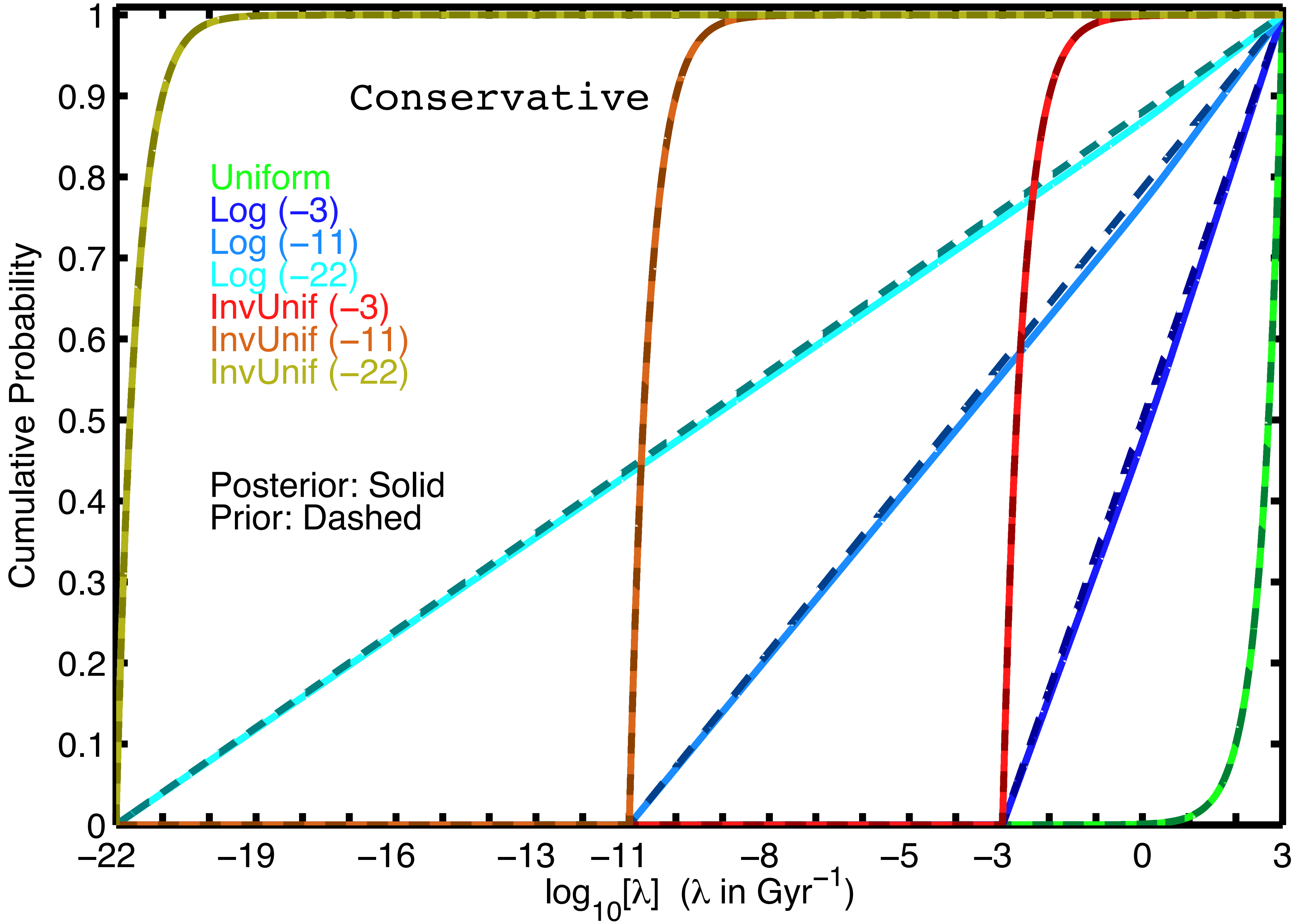
Optimistic
Uniform
Log (−3)
Log (−11)
Log (−22)
InvUnif (−3)
InvUnif (−11)
InvUnif (−22)
Posterior: Solid
Prior: Dashed
Cumulative Probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
−22
−19
−16
−13
−11
−8
−5
−3
0
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})



Median Value of λ
1−σ Lower Bound
2−σ Lower Bound
Optimistic
(λmax = 10^3 Gyr^−1)
Bounds on log10[λ] (λ in Gyr^−1)
log10[λmin] (λ in Gyr^−1)



Hypothetical
Uniform
Log (−3)
Log (−11)
Log (−22)
InvUnif (−3)
InvUnif (−11)
InvUnif (−22)
Posterior: Solid
Prior: Dashed
Cumulative Probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
−22
−19
−16
−13
−11
−8
−5
−3
0
3
log10[λ] (λ in Gyr−1)


Conservative
Uniform
Log (−3)
Log (−11)
Log (−22)
InvUnif (−3)
InvUnif (−11)
InvUnif (−22)
Posterior: Solid
Prior: Dashed
Cumulative Probability
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
−22
−19
−16
−13
−11
−8
−5
−3
0
3
log$_{10}$[λ] (λ in Gyr^{-1})

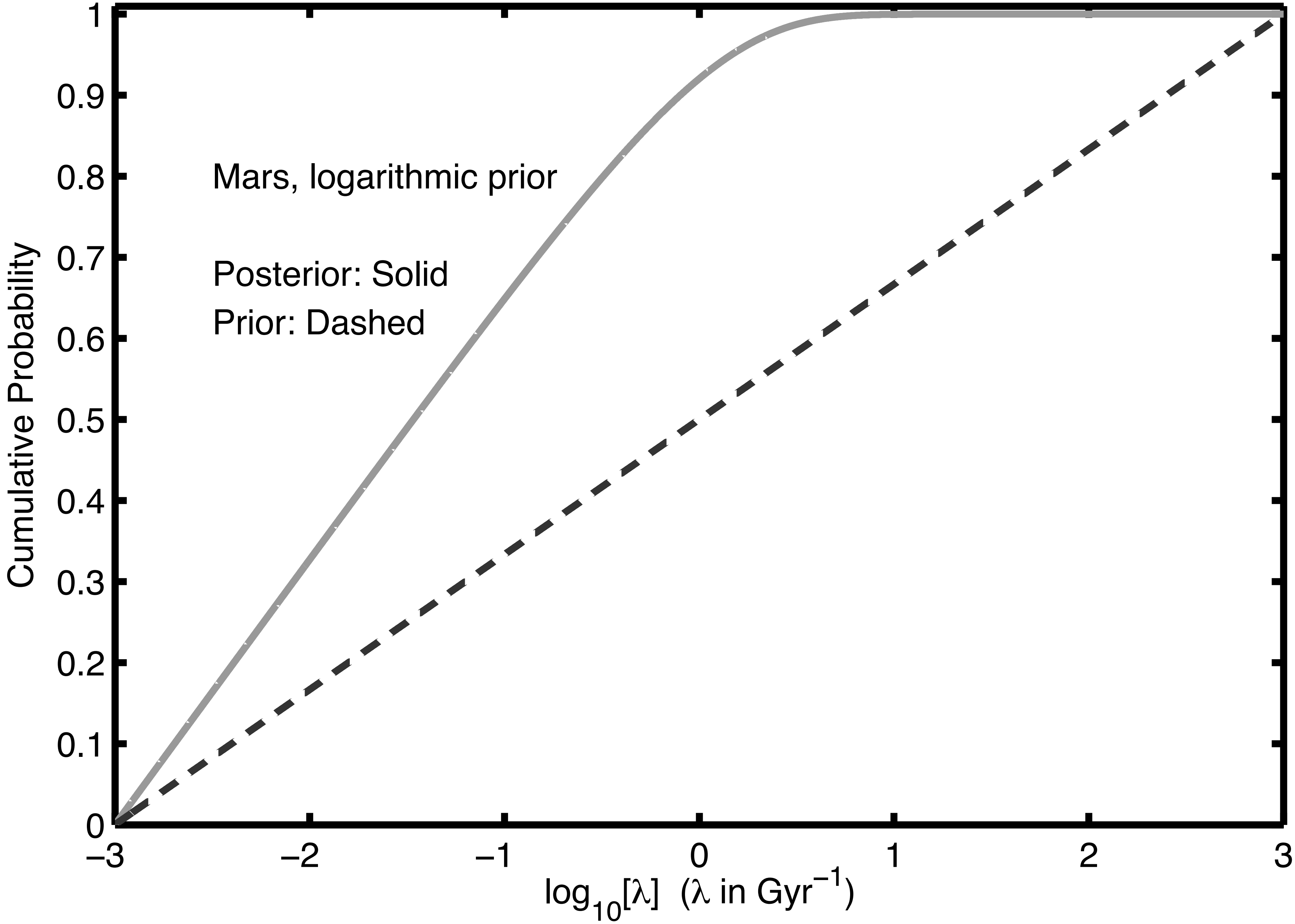
Summary

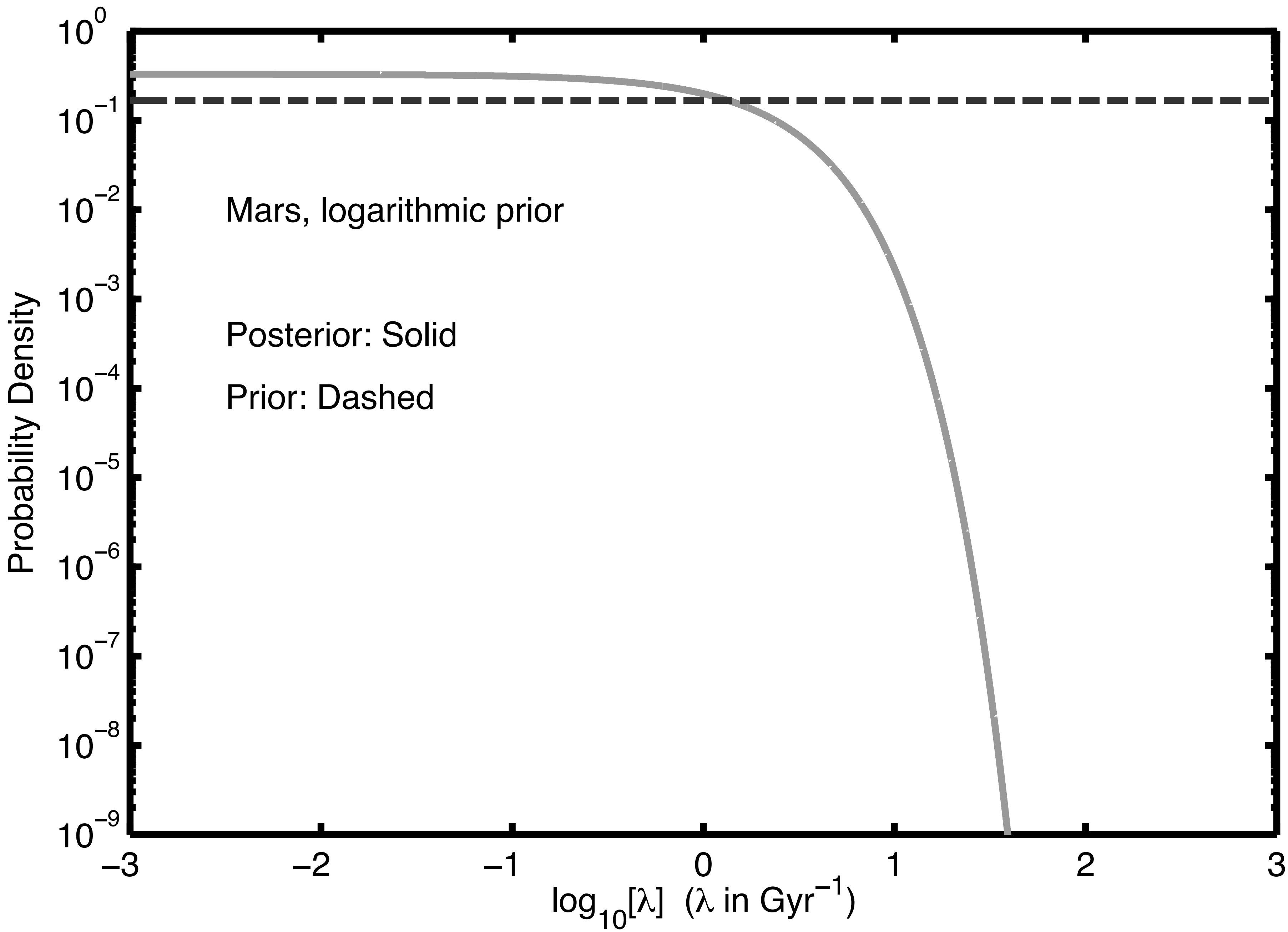
- Paleobiological & geological evidence favors a very fast development of life on the early Earth.
- If so, this datum implies a best estimate value of λ very roughly of order ~1 Gyr^{-1} or greater.
- However, even *extremely* small values of λ, that nearly exclude life elsewhere in the observable universe, are consistent with the datum.
- Stronger intuitive or formal conclusions about λ are the consequence of an informative prior and <u>not an implication of the datum.</u>

Conclusions

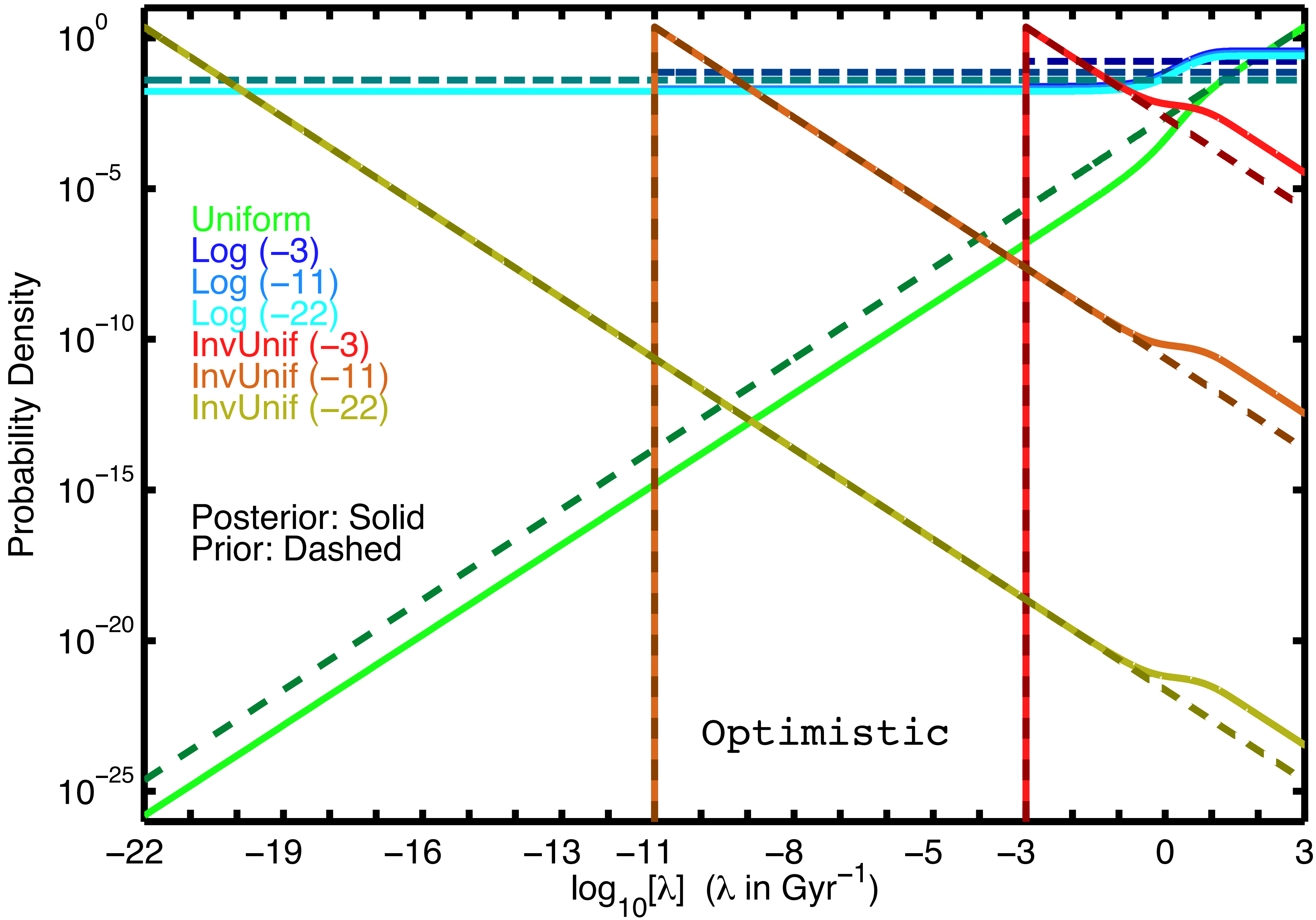
A Bayesian fan of extraterrestrial life should be encouraged by, but not highly confident based on, the rapid emergence of life on the early Earth.

A straightforward, but careful, Bayesian analysis can yield transparent and well founded results even in situations involving (very) small number statistics, anthropic considerations, intuitive arguments, uncertain data etc.

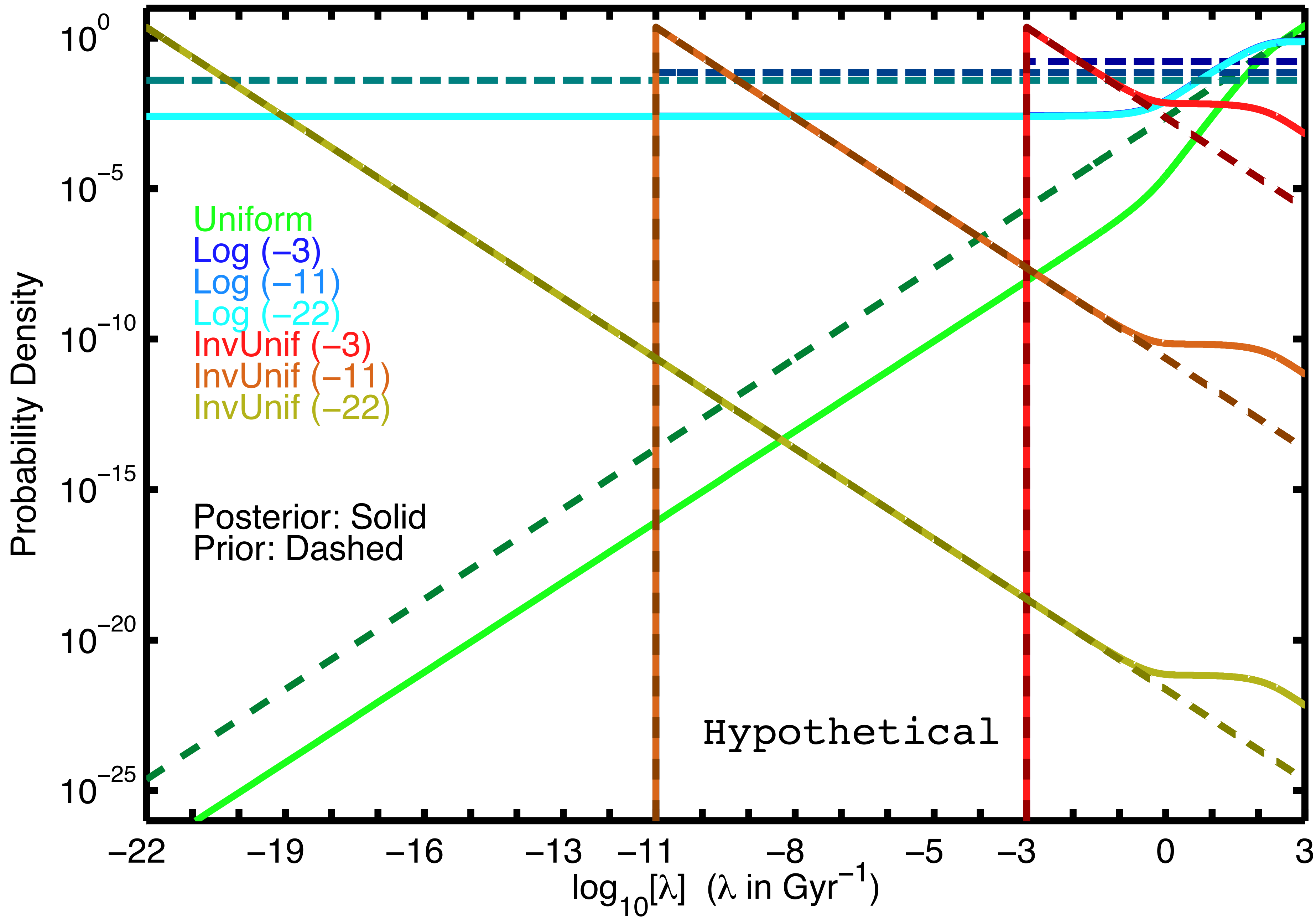

Mars, logarithmic prior
Posterior: Solid
Prior: Dashed
Cumulative Probability
log10[λ] (λ in Gyr−1)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
−3
−2
−1
0
1
2
3



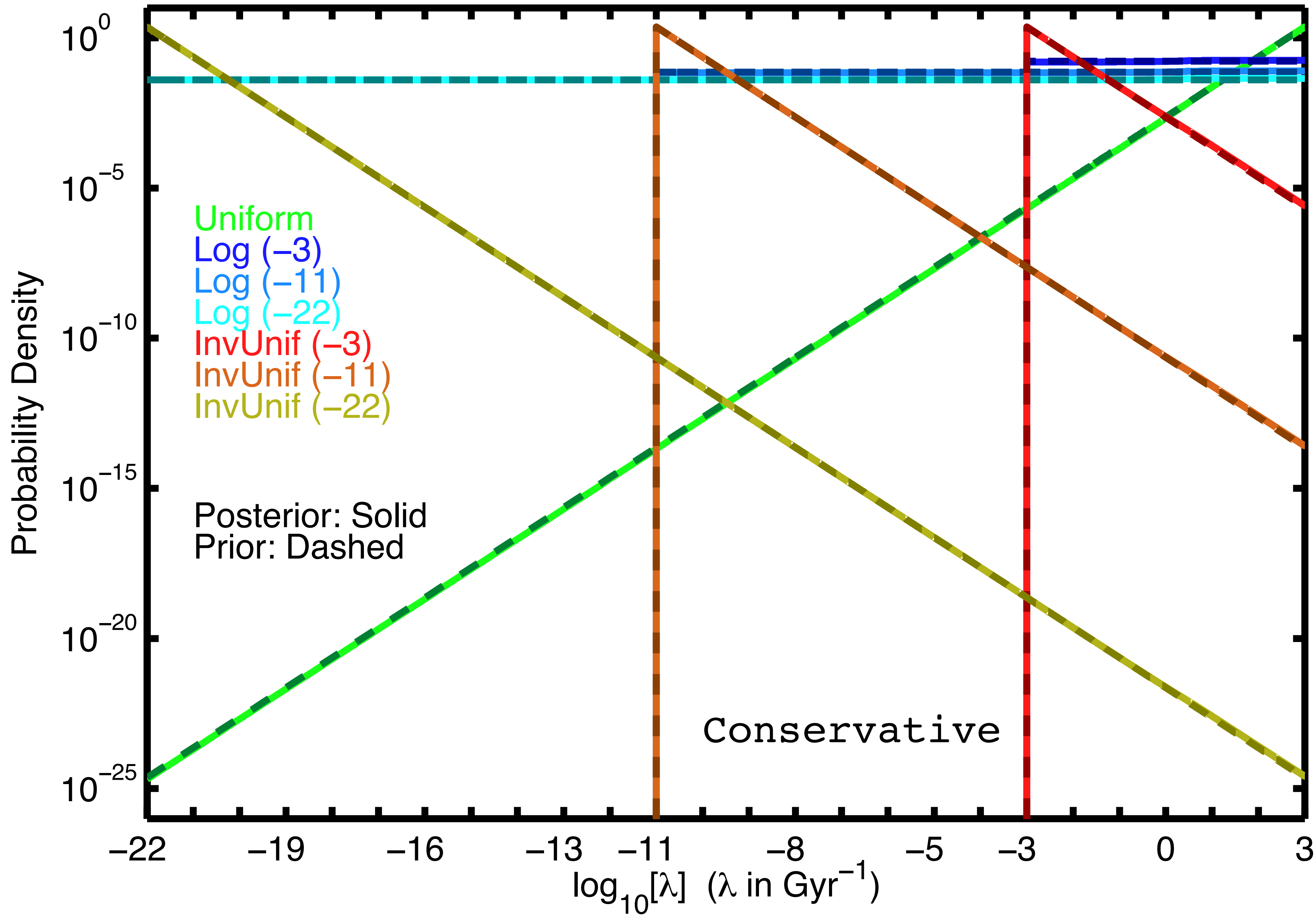
Mars, logarithmic prior
Posterior: Solid
Prior: Dashed
Probability Density
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})



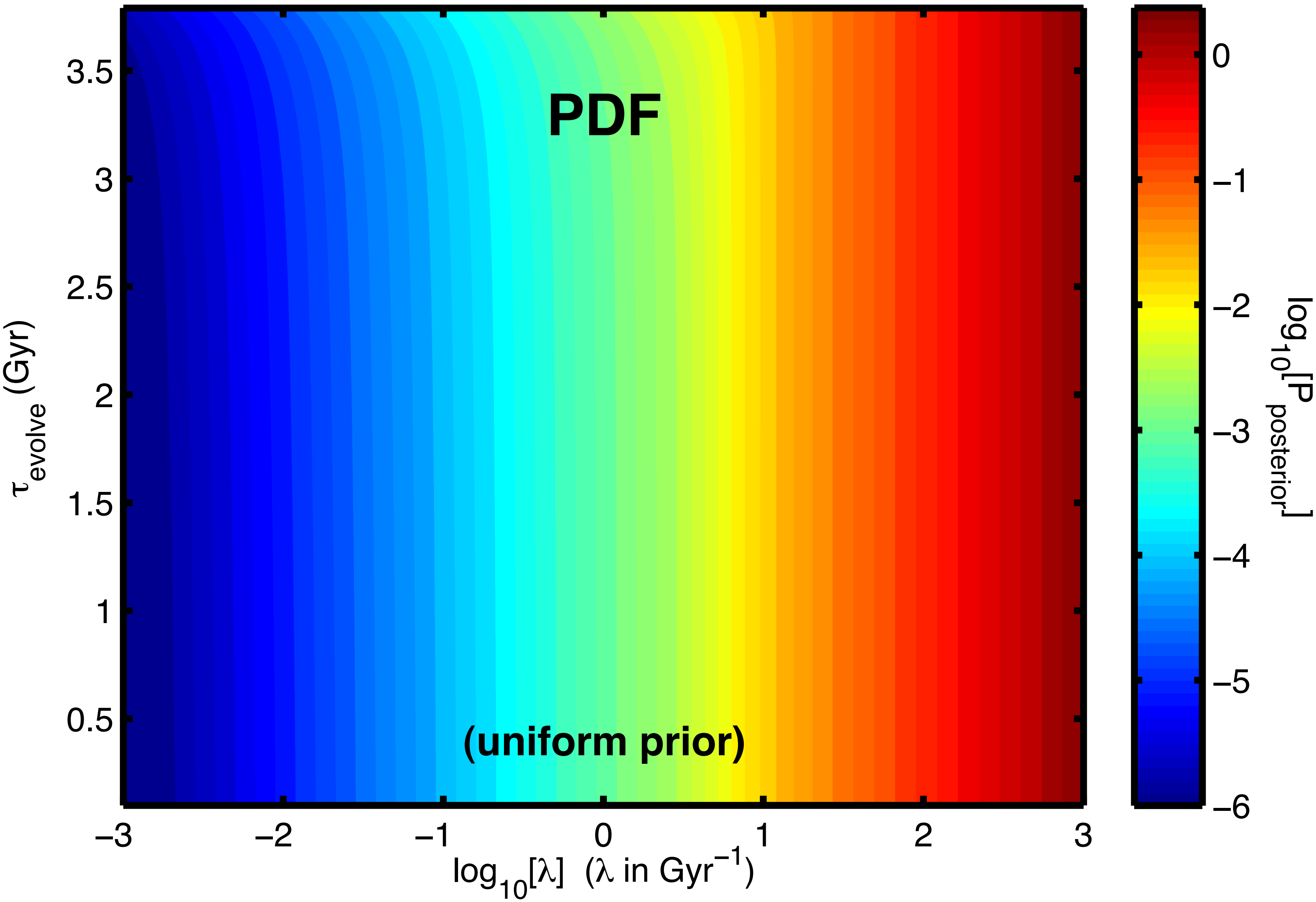
Uniform
Log (−3)
Log (−11)
Log (−22)
InvUnif (−3)
InvUnif (−11)
InvUnif (−22)
Posterior: Solid
Prior: Dashed
Optimistic
Probability Density
10^{0}
10^{-5}
10^{-10}
10^{-15}
10^{-20}
10^{-25}
−22
−19
−16
−13
−11
−8
−5
−3
0
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})

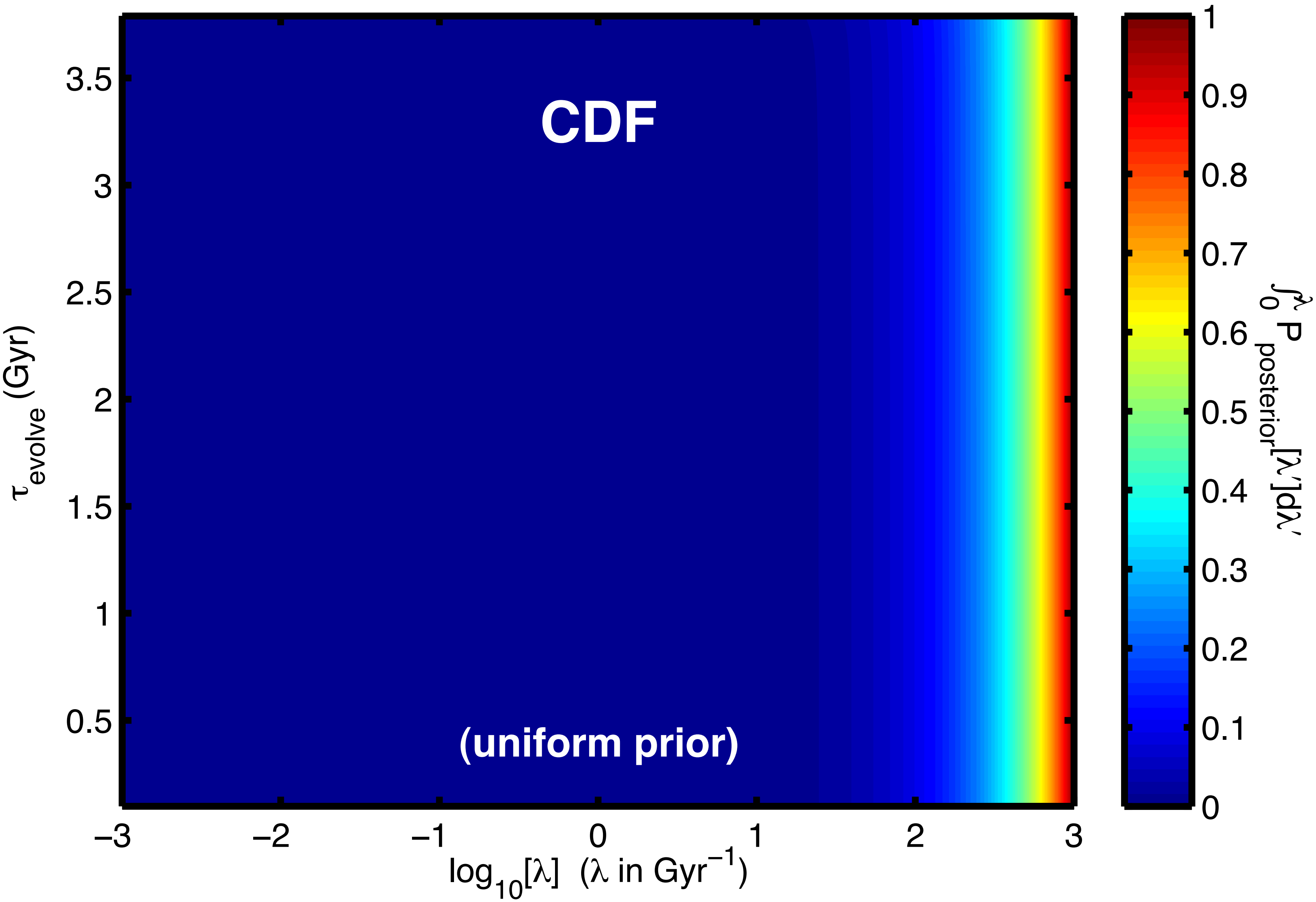


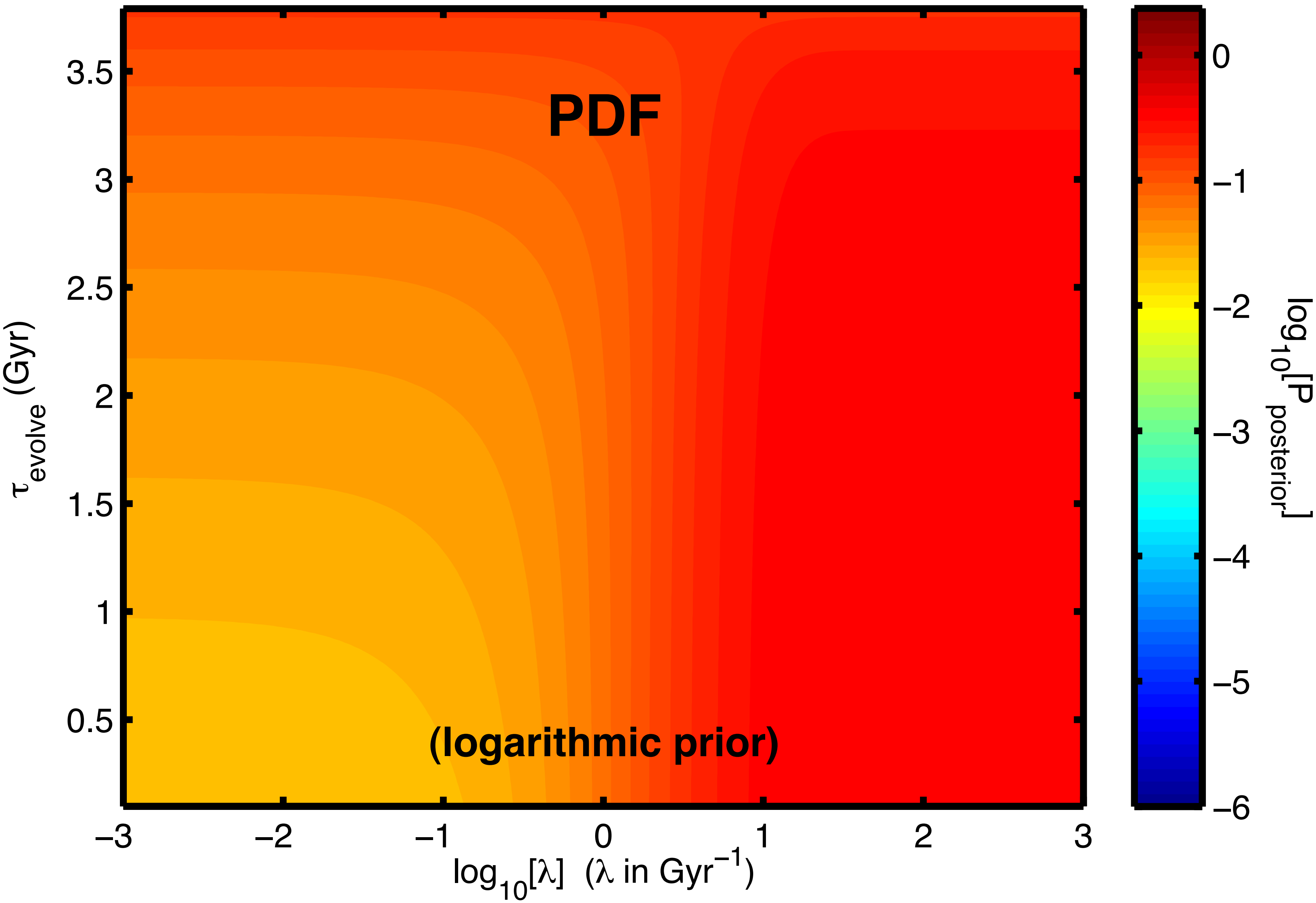
Probability Density
10^{0}
10^{-5}
10^{-10}
10^{-15}
10^{-20}
10^{-25}
Uniform
Log (−3)
Log (−11)
Log (−22)
InvUnif (−3)
InvUnif (−11)
InvUnif (−22)
Posterior: Solid
Prior: Dashed
Hypothetical
−22
−19
−16
−13
−11
−8
−5
−3
0
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})

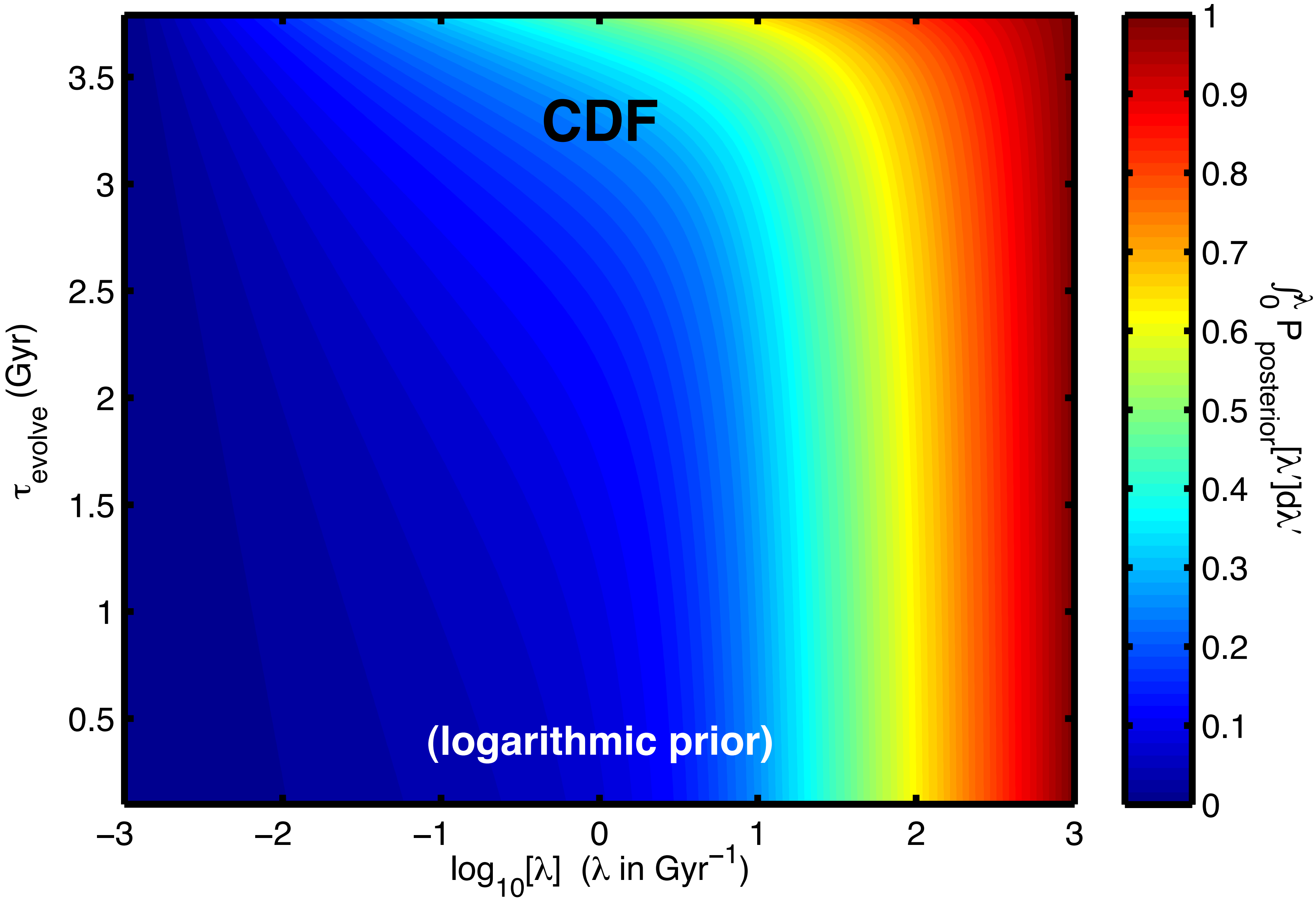


Uniform
Log (−3)
Log (−11)
Log (−22)
InvUnif (−3)
InvUnif (−11)
InvUnif (−22)
Posterior: Solid
Prior: Dashed
Conservative
Probability Density
10^{0}
10^{-5}
10^{-10}
10^{-15}
10^{-20}
10^{-25}
−22
−19
−16
−13
−11
−8
−5
−3
0
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})

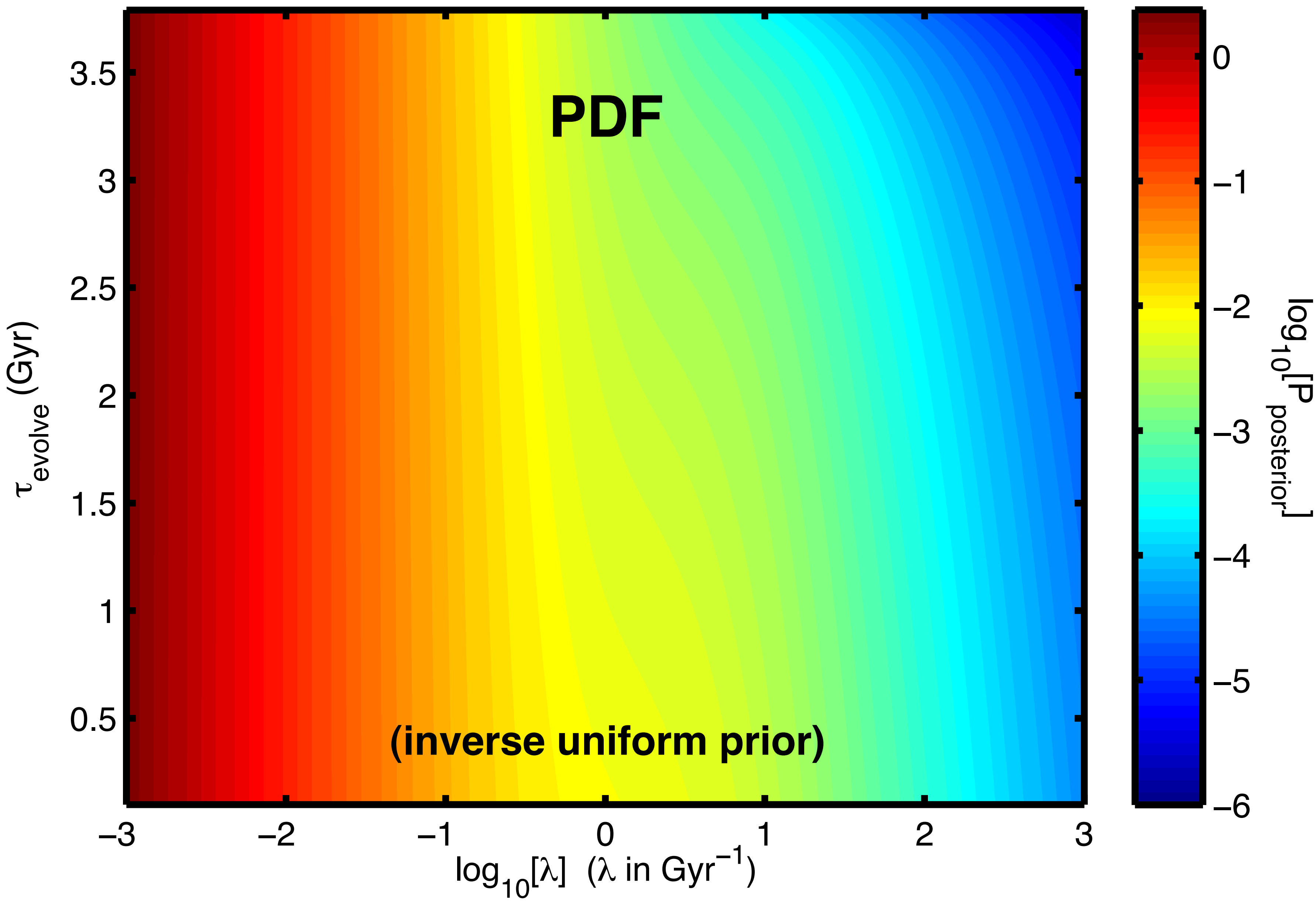

PDF
(uniform prior)
τ_{evolve} (Gyr)
3.5
3
2.5
2
1.5
1
0.5
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
0
−1
−2
−3
−4
−5
−6
$\log_{10}[P_{posterior}]$


CDF
(uniform prior)
3.5
3
2.5
2
1.5
1
0.5
τ_{evolve} (Gyr)
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$\int_0^{\lambda} P_{posterior}[\lambda']d\lambda'$

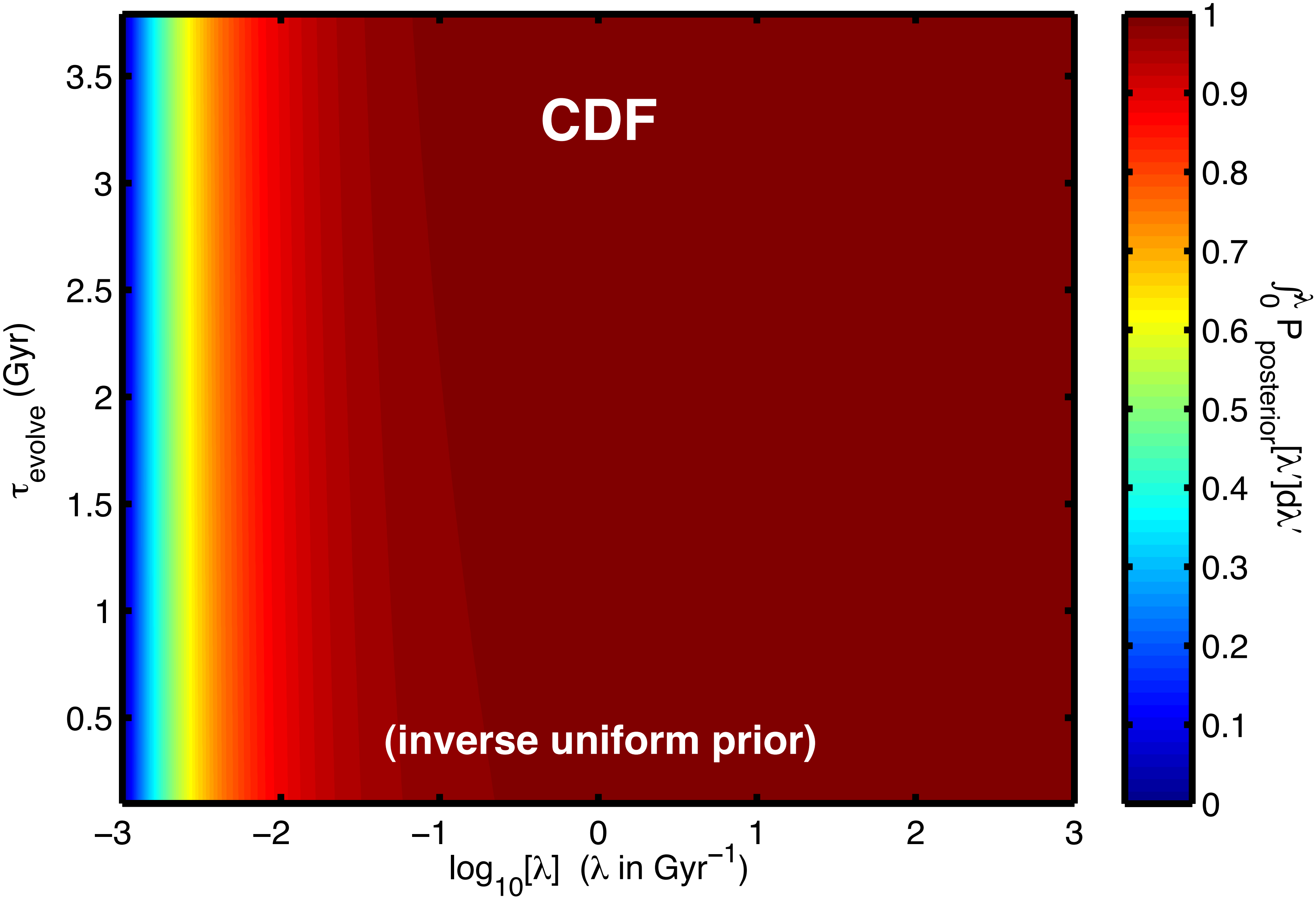

PDF
(logarithmic prior)
τ_{evolve} (Gyr)
3.5
3
2.5
2
1.5
1
0.5
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
0
−1
−2
−3
−4
−5
−6
$\log_{10}[P_{posterior}]$



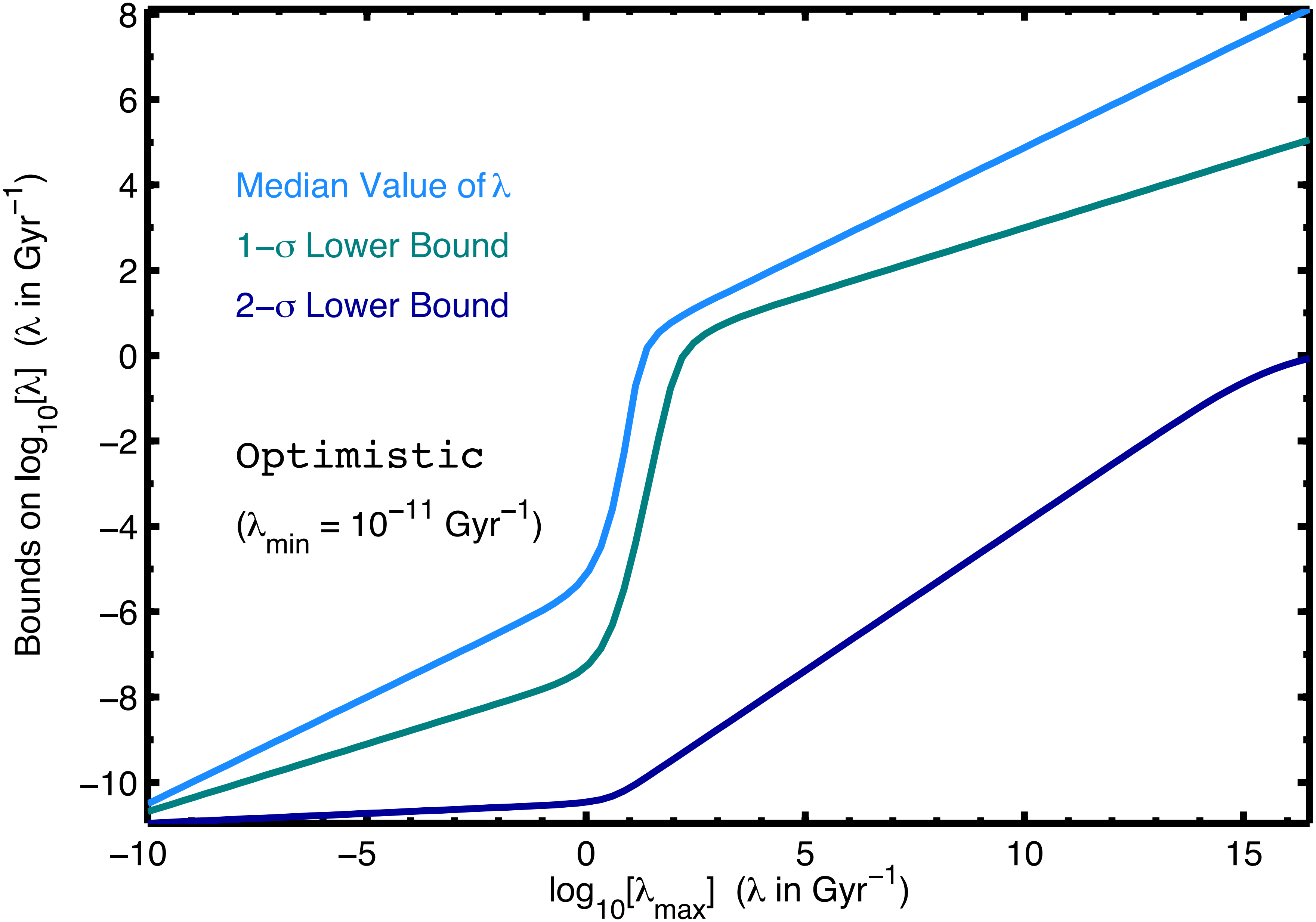
CDF
(logarithmic prior)
3.5
3
2.5
2
1.5
1
0.5
τ_{evolve} (Gyr)
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$\int_0^{\lambda} P_{posterior}[\lambda']d\lambda'$



PDF
(inverse uniform prior)
τ_{evolve} (Gyr)
0.5
1
1.5
2
2.5
3
3.5
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
$\log_{10}[P_{posterior}]$
0
−1
−2
−3
−4
−5
−6



CDF
(inverse uniform prior)
τ_{evolve} (Gyr)
3.5
3
2.5
2
1.5
1
0.5
−3
−2
−1
0
1
2
3
$\log_{10}[\lambda]$ (λ in Gyr^{-1})
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
$\int_0^{\lambda} P_{posterior}[\lambda']d\lambda'$



Median Value of λ
1–σ Lower Bound
2–σ Lower Bound
Optimistic
($\lambda_{min} = 10^{-11}$ Gyr^{-1})
Bounds on $\log_{10}[\lambda]$ (λ in Gyr^{-1})
$\log_{10}[\lambda_{max}]$ (λ in Gyr^{-1})
8
6
4
2
0
−2
−4
−6
−8
−10
−10
−5
0
5
10
15