nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0625-0

In the format provided by the authors and unedited.

Bacterial diversification through geological time

Stilianos Louca[1,2*], Patrick M. Shih[3,4,5], Matthew W. Pennell[1,2], Woodward W. Fischer[6], Laura Wegener Parfrey[1,2,7] and Michael Doebeli[1,2,8]

[1]Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada. [2]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada. [3]Joint BioEnergy Institute, Emeryville, CA, USA. [4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Department of Plant Biology, University of California, Davis, Davis, CA, USA. [6]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [7]Department of Botany, University of British Columbia, Vancouver, Brisitsh Columbia, Canada. [8]Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada.
*e-mail: louca.research@gmail.com

S.1 Mathematical derivations

This section presents mathematical formulas for the behavior of speciation-extinction cladogenic models in the limit of infinitely large bifurcating timetrees. Using simulations, we found that our mathematical formulas are accurate for trees with more than 500 tips (see Methods and Supplement S.2), as is the case for all trees examined in this study. We note that for special cases, some of the presented formulas have been discussed previously by other authors [1–3], however below we provide the full derivations for completeness. We then explain how these formulas can be used to reconstruct past diversification dynamics using trees comprising only extant OTUs. We will use the term "OTU" to refer to any group of closely related extant organisms represented as an individual tip in the tree.

S.1.1 Cladogenesis in infinitely large trees

Let T be the time point at which taxa are sampled to construct the considered tree, that is, today. We will focus on the total number of extant lineages ("total diversity", denoted $N(t)$) at each time point t as well as the number of lineages represented in the considered tree, that is, with at least one extant discovered representative at time T ("lineages through time" or LTT, denoted $\tilde{N}(t,T)$). We denote by $\lambda(t)$ the (per-lineage) speciation rate and by $\mu(t)$ the (per-lineage) extinction rate at any time t. We assume that speciations and extinctions occur randomly across the tree at Poissonian rates (i.e., with exponential uncorrelated waiting times), with all extant lineages being equally likely to speciate and equally likely to go extinct (but see our discussion on heterogenous rates in Supplement S.5). For any time t in the past ($t \leq T$), denote by $E(t,T)$ the probability that an extant lineage at time t is absent from the tree at time T, that is, has no extant discovered descendant at time T. In other words, $E(t,T)$ is the probability that the entire clade, descending from a single lineage at time t, has gone extinct at time T. Thus, $E(t,T)$ takes into account all possible speciation/extinction scenarios within that descending clade. For large trees the represented number of lineages, $\tilde{N}(t,T)$ is related to the total number of lineages, $N(t)$, as follows:

$$\tilde{N}(t,T) = N(t) \cdot [1 - E(t,T)]\,. \tag{1}$$

Reciprocally, if one knows $\tilde{N}(t,T)$ and $E(t,T)$ at some time t, one can estimate the total number of lineages $N(t)$. In the next steps, we will derive differential equations that can be used to calculate $E(t,T)$ and $N(t)$ over time, provided certain information is available about speciation and extinction rates. For trees covering only a subset of total extant OTUs (incomplete OTU sampling) we assume random and phylogenetically uncorrelated sampling, that is, we assume that extant OTUs are included or excluded from the tree independently of one another. In this case, incomplete OTU sampling at time T can be mathematically treated as an instantaneous extinction event just prior to time T that affects all lineages independently and with equal probabilities [4–6]. We later assess the accuracy of the assumption of random OTU sampling, and the effects that non-random sampling may have on the representation of past lineages in the tree (see Methods).

Consider some focal clade of size n (i.e., comprising n extant lineages) at some time point t. The transition probability rate to any other size k, denoted $\mathbb{Q}_{kn}$, depends on the speciation and extinction rates at the time, as follows:

$$\begin{aligned}
&\mathbb{Q}_{kn}(t) = n\lambda(t), \quad k = n+1,\\
&\mathbb{Q}_{kn}(t) = n\mu(t), \quad k = n-1,\\
&\mathbb{Q}_{nn}(t) = -\mathbb{Q}_{n+1,n}(t) - \mathbb{Q}_{n-1,n}(t),\\
&\mathbb{Q}_{kn}(t) = 0 \quad \text{for all other } k.
\end{aligned} \tag{2}$$

Note that $\mathbb{Q}_{kn}$ defines the transition rate matrix of a continuous-time Markov chain, whose value corresponds to the size of the focal clade over time. Hence, for a clade of size n at some time point t, the probability that the clade has size k at time $T \geq t$ is given by the (k, n)-th entry of the transition matrix, $\mathbb{M}(t, T)$, defined as the solution to the ordinary differential equation (ODE):

$$\frac{d\mathbb{M}(t,T)}{dT} = \mathbb{Q}(T) \cdot \mathbb{M}(t,T), \quad \mathbb{M}(t,t) = \mathrm{Id}, \tag{3}$$

where Id is the identity matrix. In particular, for a single extant lineage at time t (clade of size $n = 1$), the probability of extinction $E(t, T)$ is given by the $(0, 1)$-th entry (1st row, 2nd column) of the matrix $\mathbb{M}(t, T)$. The above ODE describes how $\mathbb{M}(t, T)$ evolves forward in time (i.e. for increasing T), however an analogous ODE also exists in the backward direction (i.e. for decreasing t). Indeed, according to the Chapman-Kolmogorov equation for Markov chains[7] one has

$$\begin{aligned}
\mathbb{M}(t,T) &= \mathbb{M}(t+\varepsilon,T) \cdot \mathbb{M}(t,t+\varepsilon) \\
&\overset{\text{Eq. (3)}}{=} \left[\mathbb{M}(t,T) + \varepsilon\frac{d\mathbb{M}(t,T)}{dt} + \mathcal{O}(\varepsilon^2)\right] \cdot \left[\mathrm{Id} + \varepsilon\mathbb{Q}(t) + \mathcal{O}(\varepsilon^2)\right] \\
&= \mathbb{M}(t,T) + \varepsilon\mathbb{M}(t,T)\mathbb{Q}(t) + \varepsilon\frac{d\mathbb{M}(t,T)}{dt} + \mathcal{O}(\varepsilon^2),
\end{aligned} \tag{4}$$

for any $\varepsilon \to 0$. Thus:

$$\frac{d\mathbb{M}(t,T)}{dt} = -\mathbb{M}(t,T) \cdot \mathbb{Q}(t). \tag{5}$$

In particular, since $E(t, T) = \mathbb{M}_{01}(t, T)$, one has:

$$\frac{dE(t,T)}{dt} = -\left[\mathbb{M}(t,T)\mathbb{Q}(t)\right]_{01} = -\underbrace{\mathbb{M}_{00}(t,T)}_{1}\underbrace{\mathbb{Q}_{01}(t)}_{\mu} - \underbrace{\mathbb{M}_{01}(t,T)}_{E(t,T)}\underbrace{\mathbb{Q}_{11}(t)}_{-\lambda-\mu} - \underbrace{\mathbb{M}_{02}(t,T)}_{E(t,T)^2}\underbrace{\mathbb{Q}_{21}(t)}_{\lambda}, \tag{6}$$

that is,

$$\frac{dE(t,T)}{dt} = -\mu(t) + E(t,T)\big[\lambda(t) + \mu(t)\big] - E(t,T)^2\lambda(t). \tag{7}$$

Note that we used the relationship $\mathbb{M}_{02}(t, T) = E(t, T)^2$, which holds under the assumption of this model that individual lineages survive or go extinct independently of one another.

In many cases (such as this study) it is useful to estimate total diversities in reverse time, that is, starting with the known current total number of lineages $N(T)$ and moving backward in time to estimate $N(t)$ for $t < T$. For that purpose, we re-parameterize all time functions in terms of the "age" $\tau = T - t$, for example writing $E(\tau)$ instead of $E(T - \tau, T)$. Written in terms of age, ODE (7) becomes

$$\frac{dE(\tau)}{d\tau} = \mu(\tau) - E(\tau) \cdot [\lambda(\tau) + \mu(\tau)] + E(\tau)^2\lambda(\tau). \tag{8}$$

Hence, $E(\tau)$ can be directly calculated backwards in time, assuming that $E(0)$ is known and $\lambda(\sigma)$ and $\mu(\sigma)$ are known for all younger ages $\sigma \leq \tau$. The initial value $E(0)$ corresponds to the probability of any given currently extant lineage being absent from the tree due to incomplete taxon sampling. For trees comprising all extant lineages, $E(0)$ will be zero, but for trees only comprising a random fraction ρ ("sampling fraction") of extant lineages, $E(0)$ will be equal to $1 - \rho$. Solving Eq. (8) with initial condition $E(0) = 0$ even if $\rho \neq 1$,

will yield the probability of lineage extinction regardless of whether a lineage is sampled at time T or not.

For comparison with existing common models, we briefly consider the case where λ and μ are constant over time. In that case, ODE (8) can be solved explicitly for any initial condition $E(0) = 1 - \rho$, yielding:

$$\begin{aligned} E(\tau) &= 1 - \frac{(\lambda-\mu)\rho}{\lambda\rho + [(1-\rho)\lambda - \mu]\, e^{-\tau(\lambda-\mu)}}, \quad \text{if } \lambda \neq \mu, \\ E(\tau) &= 1 - \frac{\rho}{1+\rho\tau\lambda}, \quad \text{if } \lambda = \mu. \end{aligned} \tag{9}$$

The solution in Eq. (9) has been previously derived by Nee *et al.*[2]. It is easy to see that $E(\tau)$ converges to the value μ/λ if $\mu \leq \lambda$ and to the value 1 if $\mu > \lambda$, as $\tau \to \infty$. In particular, if $\mu < \lambda$ (exponential diversification with constant speciation/extinction rates), one has $\tilde{N}(\tau) \approx (1 - \mu/\lambda)N(\tau)$ for sufficiently large τ. Hence, the ultimate probability of a lineage becoming extinct in the long run is given by μ/λ, a well known result in paleobiology[8]. Further, in the distant past $\tilde{N}$ will appear to increase at the same exponential rate as N, a result also known from previous studies[6].

The ODE in Eq. (8) further allows us to derive an ODE for the represented number of lineages $\tilde{N}(\tau)$ in the limit of infinitely large trees, as follows:

$$\begin{aligned} \frac{d\tilde{N}}{d\tau} &= \frac{d}{d\tau}\left[(1-E)N\right] \\ &= \frac{dN}{d\tau}(1-E) - N\frac{dE}{d\tau} \\ &\overset{\text{Eq. (8)}}{=} N\cdot(\mu-\lambda)\,(1-E) - N\cdot\left[\mu - E(\lambda+\mu) + E^2\lambda\right] \\ &= \lambda N\cdot(2E - 1 - E^2), \end{aligned} \tag{10}$$

where we used the fact that $dN/d\tau = (\mu - \lambda)N$. Evaluating Eq. (10) at age $\tau = 0$ yields:

$$\frac{1}{\tilde{N}}\frac{d\tilde{N}}{d\tau}\Big|_{\tau=0} = \lambda(0)\frac{N(0)}{\tilde{N}(0)}\left[2(1-\rho) - 1 - (1-\rho)^2\right] = -\rho\lambda(0). \tag{11}$$

In particular, in the case of complete sampling ($\rho = 1$) at $\tau = 0$ one has $\tilde{N}^{-1}d\tilde{N}/d\tau = -\lambda$, that is, at the tips of the tree the exponential growth rate of $\tilde{N}$ is equal to the speciation rate. This is a well-known result in phylogenetics[6], which is commonly explained by the fact that recently emerged lineages have not had the time yet to go extinct, making the tree's recent growth appear as if extinction rates were zero.

S.1.2 Fitting speciation/extinction models to trees

If the absolute speciation and extinction rates $B = \lambda N$ and $D = \mu N$ are known, then integration of the ODE

$$\frac{dN}{d\tau} = D(\tau) - B(\tau) \tag{12}$$

allows calculation of the total diversity N over time. Further, knowledge of the sampling fraction ρ allows the calculation of $E(\tau)$ by integrating ODE (8) backwards in time, and therefore of $\tilde{N}(\tau)$ via Eq. (1). Today's total number of extant lineages, $N(0) = \tilde{N}(0)/\rho$, can be used as an initial condition for ODE (12). Thus, for a given tree, a given sampling fraction ρ and a model that specifies B and D as functions of τ and/or

N, the LTT predicted by the model, $\tilde{N}_m(\tau)$, can be compared to the LTT observed from the tree, $\tilde{N}(\tau)$. In principle, then, it is possible to fit model parameters by minimizing the deviation of the model's predicted $\tilde{N}_m$ from the observed $\tilde{N}$.

In the simplest case the (per-lineage) speciation and extinction rates may be assumed to be constant over time, i.e. $B = \lambda N$ and $D = \mu N$ for constant parameters λ, μ. In general, however, speciation and/or extinction rates could themselves depend on current diversities or depend explicitly on time, and thus additional model parameters may be needed to describe these dependencies[5]. As we explain in the main text, in this study we found that constant speciation/extinction rate models provide adequate descriptions of global bacterial and cyanobacterial diversification dynamics at the temporal resolutions and time intervals considered (mean relative deviation below 5% in all cases).

S.1.3 Non-parametric estimations

Certain useful quantities may also be estimated non-parametrically from the LTT, that is, using the value and derivatives of the LTT at each time point, instead of fitting a parametric model to the entire LTT curve. The advantage of non-parametric estimations over parametric models is that they make fewer or no assumptions on how speciation and extinction rates vary over time, and they do not include a tradeoff between temporal resolution and model simplicity (e.g., in terms of the number of allowed rate shifts). Below, we describe novel non-parametric methods for extracting information from LTTs. In Supplement S.2 we demonstrate the power of these methods, using trees simulated under various scenarios.

We denote by $\tilde{\nu} = (1/\tilde{N})d\tilde{N}/d\tau$ the relative slope of the LTT. Note that $\tilde{\nu}$ is the apparent diversification rate that would be measured solely based on the LTT without consideration of extinctions or incomplete taxon sampling. As we show below, $\tilde{\nu}$ and its slope at any time in the past can yield important information on speciation and extinction rates at that time. We note that examining $\tilde{\nu}$ is analogous to examining branching frequencies in the tree over time, since $\tilde{\nu}$ is proportional to the branching frequency per lineage. Dividing Eq. (10) by Eq. (1), yields the relationship

$$\tilde{\nu} = \frac{1}{\tilde{N}}\frac{d\tilde{N}}{d\tau} = \lambda \cdot (E - 1). \tag{13}$$

Solving Eq. (13) for E yields

$$E(\tau) = 1 + \frac{\tilde{\nu}(\tau)}{\lambda(\tau)}. \tag{14}$$

Inserting Eq. (14) into Eq. (1) yields an expression for the total diversity:

$$\boxed{N(\tau) = \frac{\tilde{N}(\tau)}{1 - E(\tau)} = -\frac{\lambda(\tau)\tilde{N}(\tau)}{\tilde{\nu}(\tau)}.} \tag{15}$$

Hence, if the speciation rate $\lambda(\tau)$ is known at some age τ, one can estimate $E(\tau)$ and $N(\tau)$ directly from the LTT using Eq. (14) and (15), respectively. Subsequently, one may estimate $\mu(\tau)$ based on the relationship:

$$\frac{dN}{d\tau} = -\frac{dN}{dt} = -r \cdot N = -(\lambda - \mu) \cdot N, \tag{16}$$

where $r = \lambda - \mu$ is the diversification rate. From Eq. (16) one obtains:

$$\mu = \lambda + \frac{1}{N}\frac{dN}{d\tau}. \tag{17}$$

We emphasize that the speciation rate $\lambda(\tau)$ is usually unknown except for $\tau = 0$, where $\lambda(\tau = 0) = -\tilde{\nu}(\tau = 0)/\rho$ according to Eq. (11). In the special case where λ is approximately constant over time, λ can be estimated as $-\tilde{\nu}(0)/\rho$ or estimated by fitting a speciation-extinction model (as done in this study), and subsequently E, N and μ can be estimated from the LTT using Eqs. (14), (15) and (17). If λ varies over time, the erroneous assumption of a constant λ, equal to $\lambda(0)$, would introduce an error into the estimated E, N and μ. In that case, it may be useful to consider modified quantities, described below, that can be estimated from the LTT regardless of whether λ is constant or not.

Pulled total diversity

The first modified quantity, which we refer to as "*pulled total diversity*", is defined as:

$$\boxed{N_{\mathrm{p}}(\tau) = N(\tau) \cdot \frac{\lambda_o}{\lambda(\tau)},} \tag{18}$$

where $\lambda_o = \lambda(0)$. Observe that N_{p} is similar to the total diversity N, but differs from the latter by the factor $\lambda(0)/\lambda(\tau)$. From Eq. (15) it becomes clear that

$$N_{\mathrm{p}}(\tau) = -\frac{\lambda_o \tilde{N}(\tau)}{\tilde{\nu}(\tau)}. \tag{19}$$

Since λ_o can be estimated from the LTT near the tips of the tree, the right hand side of Eq. (19) can be estimated directly from the LTT without any *a priori* assumptions on how λ may have varied over time. If λ is approximately constant over time, then $N_{\mathrm{p}}(\tau)$ will be approximately equal to $N(\tau)$. In particular, unless λ changed drastically (i.e., by orders of magnitude) over time, N_{p} provides a quick way to estimate past total diversities to order of magnitude accuracy.

Pulled extinction rate

The second modified quantity, which we refer to as "*pulled extinction rate*", is defined as:

$$\mu_{\mathrm{p}}(\tau) = \mu(\tau) + [\lambda_o - \lambda(\tau)] - \frac{1}{\lambda}\frac{d\lambda}{d\tau}. \tag{20}$$

Observe that μ_{p} is similar to the extinction rate μ, but differs from the latter by the terms $[\lambda_o - \lambda(\tau)]$ and $\lambda^{-1} d\lambda/d\tau$, both of which disappear when λ is constant over time. Eq. (20) can be rewritten as

$$\mu_{\mathrm{p}}(\tau) = \lambda_o - r(t) - \frac{1}{\lambda}\frac{d\lambda}{d\tau}. \tag{21}$$

Hence, if diversity is close to speciation/extinction equilibrium ($|r| \ll \lambda_o$), and λ changes only slowly, the pulled extinction rate is approximately equal to the recent extinction rate λ_o. Using Eqs. (15) and (17) in Eq.

(20) yields the formula:

$$\boxed{\mu_{\mathrm{p}}(\tau) = \lambda_o + \frac{d}{d\tau} \ln \left[-\frac{\tilde{N}(\tau)}{\tilde{\nu}(\tau)} \right].} \tag{22}$$

Similarly to the pulled total diversity, the pulled extinction rate can be estimated from the LTT without any *a priori* assumptions on how λ and μ may have varied over time.

Pulled diversification rate

To derive the third modified quantity, we insert Eq. (15) into Eq. (17), thus obtaining the relationship:

$$\lambda - \mu + \frac{1}{\lambda}\frac{d\lambda}{d\tau} = -\frac{d}{d\tau} \ln \left[-\frac{\tilde{N}(\tau)}{\tilde{\nu}(\tau)} \right] = -\tilde{\nu} + \frac{1}{\tilde{\nu}}\frac{d\tilde{\nu}}{d\tau}. \tag{23}$$

Observe that the right hand side of Eq. (23) can be readily calculated from the LTT without any assumption on λ, μ or ρ. The left hand side resembles the diversification rate ($r = \lambda - \mu$), modified ("pulled") by an additive correction term. We define the left hand side as the "*pulled diversification rate*" (PDR):

$$\boxed{r_{\mathrm{p}} = r + \frac{1}{\lambda}\frac{d\lambda}{d\tau},} \tag{24}$$

so that

$$r_{\mathrm{p}}(\tau) = -\tilde{\nu} + \frac{1}{\tilde{\nu}}\frac{d\tilde{\nu}}{d\tau}. \tag{25}$$

We point out that the "pulled" variables (μ_{p}, r_{p}, N_{p}) and the recent speciation rate λ_o satisfy similar algebraic relationships as their "non-pulled" counterparts:

$$r_{\mathrm{p}} = \lambda_o - \mu_{\mathrm{p}}, \tag{26}$$

and

$$\frac{1}{N_{\mathrm{p}}}\frac{dN_{\mathrm{p}}}{d\tau} = -r_{\mathrm{p}}. \tag{27}$$

Equation (25) can be used to estimate the pulled diversification rate from the LTT curve, without knowing λ, μ or ρ, without any assumptions on how λ or μ may have varied over time, and without fitting any model parameters (see simulations in Supplementary Figs. 1–6 for examples). The downside is that, similarly to the pulled total diversity and the pulled extinction rate, r_{p} is a composite quantity that only resembles the diversification rate r when λ is roughly constant. If λ changes rapidly over time, r_{p} will differ substantially from r. Hence, solely knowing r_{p} does not *a priori* determine its constituents μ and λ, nor r, unless either λ or μ is estimated separately (or assumed to be constant).

S.2 Evaluating non-parametric methods using simulations

To demonstrate the use of non-parametric estimation methods introduced in Supplement S.1.3, and to assess how the pulled extinction rates (PERs), pulled diversification rates (PDRs) and pulled total diversities (PTDs)

differ from their "non-pulled" analogs, we tested these methods on trees simulated under various scenarios (Figs. 1 and 1–6). Specifically, we simulated various trees with constant or non-constant speciation rates, and then examined the behavior of the PERs, PDRs and PTDs. In all scenarios, a random tree was generated according to the general model discussed in Supplement S.1.1, whereby lineages speciate or go extinct randomly at exponentially distributed time intervals and independently of one another. At any moment in time, speciation and extinction rates are assumed to be the same (homogenous) across all clades, but can vary over time. For an evaluation of how rate heterogeneity across clades could affect our results, see Supplement S.5. Following a simulation, the PER (μ_{p}) was estimated based on the generated tree's LTT using formula (22), the PDR (r_{p}) was estimated using formula (25) and the PTD (N_{p}) was estimated using formula (19). Details of each simulation scenario are described below.

In the first examined scenario (Fig. 1), a tree was generated with constant speciation rate λ, and a density-dependent extinction rate ($\mu \propto N^{0.2}$) with an additional temporary increase about 80 Ma ago, simulating a mass extinction event. Note that, since λ was constant, in this scenario the PER was equal to the extinction rate, the PDR was equal to the diversification rate and the PTD was equal to the total diversity. As seen in Fig. 1a, the total diversity (or PTD) is estimated very accurately except for the very early times near the root, and clearly reveals the effects of the mass extinction event. Similarly, the PDR, as estimated using Eq. (25), closely resembles the true PDR and clearly reflects the temporary drop in the diversification rate (Fig. 1b). The PER contains substantial noise and wide confidence intervals, although the mass extinction event is again clearly reproduced.

In the second scenario (Fig. 2), a tree was generated using a constant speciation rate and an extinction rate exhibiting a sharp temporary increase (mass extinction event, lasting only $\sim$2 Ma) about 90 Ma ago. Similarly to the first scenario, the spike in the extinction rate induces a spike in the PER (Fig. 2C) and in the PDR (Fig. 2b). Similarly, the tree's LTT, the PTD and the estimated PTD all become distorted due to the mass extinction event. Due to the brevity of the event, the spikes in the estimated PER, PDR and PTD are severely damped by the noise-filter (smoothening) applied during estimation. That said, the tree's LTT still reflects the mass extinction event (Fig. 2a). In particular, fitting a constant-rates cladogenic model, as done for the bacterial trees, reveals a clear deviation from the LTT near the mass extinction event (Fig. 2d).

In the third scenario (Fig. 3), a tree was generated using a strongly oscillating speciation rate ($\pm$50% around the mean, with period 200 million years) and a constant extinction rate. The estimated PDR clearly reflects the oscillatory nature of the diversification rate (Fig. 3b), although noticeable differences exist compared to the diversification rate due to the "pulling" term $\lambda^{-1}d\lambda/d\tau$ discussed in Eq. (25). The estimated PER is strongly distorted by the oscillations in the true speciation rate (Fig. 3c). Interestingly, the estimated PTD only deviates moderately from the true total diversity, when compared on a logarithmic axis. This is because λ is always within the same order of magnitude as λ_o, and hence N_{p} also is always within the same order of magnitude as N (Eq. (18)).

In the fourth scenario (Fig. 4), a tree was generated using a constant extinction rate, and a diversity-dependent speciation rate that increased as the tree grew over time ($\lambda \propto N^{0.1}$). As a result, near the tips λ was about twice as large as near the root. The diversity-dependence-driven increase in the speciation rate leads to a positive trend in the PDR (Fig. 4b), a negative trend in the PER (Fig. 4c), and a systematic difference between N_{p} and N that increases towards older ages (Fig. 4a).

In the fifth scenario (Fig. 5), a tree was generated using a constant extinction rate, and a speciation rate exhibiting a strong temporary increase (spike) about 90 Ma ago. The variation in the speciation rate causes a strong distortion in the PER, which may be confused with a temporary drop in extinction rate (Fig. 5c). The PDR also clearly reflects the temporary increase in diversification rate, and the difference between the two is only moderate (Fig. 5b). Similarly, the PTD approximately resembles the total diversity, and clearly

reproduces the sudden increase of diversity during the speciation spike (Fig. 5a).

In the sixth scenario (Fig. 6), a tree was generated using a constant extinction rate, and a speciation rate exhibiting a strong, sharp, temporary increase about 140 Ma ago ("speciation spike", lasting only ∼1 Ma). Similarly to the previous scenario, the spike in the speciation rate induces a strong spike in the PER (Fig. 6c) and in the PDR (Fig. 6b). Similarly, the tree's LTT, the PTD and the estimated PTD all become distorted due to the speciation spike. Due to the brevity of the speciation spike, the spikes in the estimated PER, PDR and PTD are severely damped by the noise-filter (smoothening) applied during estimation. That said, the tree's LTT clearly reflects the past anomaly in the speciation rate (Fig. 6a). In particular, fitting a constant-rates cladogenic model, as done for the bacterial trees, does not accurately reproduce the tree's LTT (Fig. 6d).

The above examples demonstrate three important points: First, for large trees generated under a constant speciation rate, non-parametric estimation methods can accurately reconstruct extinction rates, diversification rates and total diversities over time. Second, if the speciation rate varies over time (for example due to diversity-dependence), then this variation generally (i.e. if μ is chosen arbitrarily) leads to similarly strong variations in the PER, even if the extinction rate is constant over time. It is generally hard (using solely the LTT) to determine whether variations in the PER and/or PDR are due to a varying speciation rate or due to a varying extinction rate. If, however, the estimated PER turns out to be constant over time, this is a strong indication that both λ and μ were both constant or varied only slowly over time (see detailed explanation in Supplement S.4). Third, the magnitude of the PDR is generally comparable to the magnitude of the diversification rate, except perhaps during short isolated time intervals. In fact in all of the above simulations the PDR approximately resembled the diversification rate.

S.3 Robustness of global extant bacterial diversity estimates

Here we discuss the robustness of our estimates of global extant bacterial OTU richness. Our estimate of 1.4 million bacterial OTUs (discussed in the main article) remains approximately the same (∼1.1 million OTUs) when we omit *de novo* OTUs found in fewer than 10 samples (68,927 OTUs kept, of which ∼42% were covered by SILVA), suggesting that a potential bias towards more ubiquitous OTUs only weakly affects global bacterial diversity estimates. Our estimate also remains approximately the same when we strongly subsample reads (keeping ∼13,000 reads per sample on average, instead of ∼40,000 in the full dataset), yielding an estimated ∼1.2 million OTUs. The weak dependence on sequencing depth shows that our approach is not substantially affected by a detection bias towards more abundant organisms in each sample. The estimate is also similar (∼1.6 million OTUs) when we consider 16S rRNA sequences from metagenome-assembled genomes[9], instead of OTUs recovered from amplicon sequences, indicating that primer bias also only has a moderate influence. Finally, our estimate is similar (∼1.9 million OTUs) when we consider the overlap between our *de novo* OTUs and OTUs that we recovered from another massive dataset generated by the Earth Microbiome Project (EMP)[10], instead of an overlap between our *de novo* OTUs and SILVA (Supplementary Table 3). Our estimate is comparable to another recent estimate (∼1.5 million unique 16S rRNA sequences)[11], but is 6 orders of magnitude lower than the estimate by Locey *et al.*[12], which was based on extrapolation of empirical scaling laws of local richness to global scales. The strong disagreement with Locey *et al.*'s estimates supports arguments that empirical scaling laws of local richness cannot be extrapolated to estimate global microbial richness[13]. Further, many of the extremely rare but diverse OTUs found in previous studies (termed the "rare biosphere"), such as in early releases of the EMP[14], which reportedly recovered millions of rare V4-OTUs, may have been spurious due to methodology. Spurious OTUs are known to frequently occur in amplicon sequencing studies and can substantially inflate microbial richness estimates, notably due to sequencing errors and chimera formation during PCR[15,16]. Indeed, a more recent

release of the EMP that employed improved quality filters found only ∼307,000 unique V4 sequences across 96 studies [10].

S.4 Interpreting pulled variables of bacterial and cyanobacterial trees

Pulled extinction rates

As discussed in the main text, here we found that estimated bacterial and cyanobacterial pulled extinction rates were almost constant over time. As we explain below, this strongly indicates that both λ and μ were constant or varied only slowly over time. To see why this is the case, observe that for any time point the right hand side in Eq. (20) depends solely on the instantaneous speciation and extinction rates at that time point. Hence, if μ and/or varied over time, $\mu_{\rm p}$ could only be constant if μ instantaneously adjusted to (was determined by) the current value of λ and its relative rate of change, or vice versa. Specifically, for any given μ, according to Eq. (21) $\mu_{\rm p}$ could only be (or estimated to be) approximately constant over time ($|\mu_{\rm p}(t) - \mu_{\rm p}(T)| < \varepsilon$ for some small ε) if $\lambda(t)$ satisfied the differential equation

$$\frac{1}{\lambda}\frac{d\lambda}{dt} \approx r(t) - r_o, \tag{28}$$

where r_o is the recent diversification rate, and the error in Eq. (28) must be smaller than ε. For strongly varying r, $\lambda^{-1}d\lambda/dt$ must have closely and instantaneously followed $r - r_o$, however there is no known realistic mechanism that would cause $\lambda^{-1}d\lambda/dt$ to behave in such way. In fact, for generic and independently chosen $\mu(t)$ and $\lambda(t)$, one or both of which exhibits variations greatly exceeding ε, Eq. (28) would almost never be satisfied. The fact that for generic μ and λ the pulled extinction rate is unlikely constant is also easily demonstrated using simulations (Figs. 1–6). It follows that $r(t)$ must have been approximately constant ($|r - r_o| \lesssim \varepsilon$) or small in magnitude ($|r| \lesssim \varepsilon$), and thus $\lambda^{-1}d\lambda/dt$ must also have been small in magnitude ($\lesssim \varepsilon$). In the case of bacteria, this means that λ must have varied at billion-year time scales ($|\lambda^{-1}d\lambda/dt| \sim 0.005\ \mathrm{Ma}^{-1}$) or slower. We note that, since λ must have been constant or varied only slowly over time, according to Eq. (20) the recent pulled extinction rate is similar to the recent extinction rate ($|\mu_{\rm p}(T) - \mu(T)| \lesssim \varepsilon$). Towards older ages (e.g., 10^3 Ma ago in the case of bacteria), $\mu_{\rm p}$ will remain close to λ_o even if λ and/or μ changed substantially (Eq. 21), as observed in Figs. 2d,e.

Pulled diversification rates

As discussed in the main text, all of our fitted models suggest low bacterial and cyanobacterial diversification rates ($r \sim 0.002$–$0.005\ \mathrm{Ma}^{-1}$, Fig. 18c). Consistent with this observation, we also find low PDRs ($|r_{\rm p}| < 0.005\ \mathrm{Ma}^{-1}$, Figs. 2e,f and 21g–i). The fact that $|r_{\rm p}|$ is small over a prolonged time interval ($|r_{\rm p}| < \varepsilon$, where $\varepsilon \sim 0.005\ \mathrm{Ma}^{-1}$) is indeed a strong indication that within that time interval the magnitude of the diversification rate ($|r|$) was also not much larger than ε. To see why this is the case, note that at each time point t the right hand side of Eq. (24) only depends on the instantaneous diversification rate and the relative rate of change of λ at that time point. For generic λ and μ with $|r| \gg \varepsilon$ one would generally (and at most time points) also expect that $|r_{\rm p}| \gg \varepsilon$, since otherwise $\lambda^{-1}d\lambda/dt$ would have to instantaneously adjust to approximately cancel out r; an unlikely coincidence. This logic is easily demonstrated using simulations, where $r_{\rm p}$ generally fluctuates at similar magnitudes as μ and/or r (see simulation examples in Figs. 1–6).

Pulled total diversities

In all examined bacterial and cyanobacterial trees, our fitted (constant-rate) models suggest a positive diversification rate and a roughly exponential increase of total diversity over the last 1 billion years, with bacterial and cyanobacterial diversity increasing roughly 10-fold (Figs. 2a,b and 21a–c). Consistent with this observation, bacterial and cyanobacterial pulled total diversities also exhibit an approximately exponential increase that closely resemble the predictions of the fitted models. An exponentially growing N_{p} suggests that N was itself growing approximately exponentially. To see why this is the case, recall that $N_{\mathrm{p}} = N\lambda/\lambda_o$ according to Eq. (18). Since $N_{\mathrm{p}}(\tau) \approx N_o e^{-r_o \tau}$, where r_o is the diversification rate fitted by the model, one has

$$e^{r_o \tau} \approx \frac{\lambda(\tau)}{N(\tau)} \cdot \frac{N_o}{\lambda_o}. \tag{29}$$

The above equation can be satisfied in 3 alternative ways. First, λ could be approximately constant ($\lambda \approx \lambda_o$), in which case N is itself growing exponentially ($N \approx N_{\mathrm{p}}$) subject to approximately constant speciation and extinction rates. Second, $N(\tau)$ may be approximately constant, in which case λ (and thus μ, since $\lambda - \mu \approx 0$) must be approximately exponentially decreasing over time. It is, however, hard to find a reasonable explanation for why λ and μ would decline exponentially at a roughly constant rate over the last 1 billion years, while N remains constant. Third, both N and λ vary substantially over time, in such a way that their ratio declines approximately exponentially. For generic varying $\lambda(t)$ and $N(t)$ this is highly unlikely, unless either λ is largely determined by N (and N varies such that $\lambda(N(\tau))/N(\tau)$ is exponentially declining), or N is largely determined by λ (and λ varies such that $\lambda(\tau)/N(\lambda(\tau))$ is exponentially declining). While the second and third scenario are mathematically perfectly valid, they are quite idiosyncratic and more complex than the first scenario, which solely requires that speciation and extinction rates be approximately constant over time. Further, both scenarios would not result in such a good fit of a constant-rates model as observed in this study (Figs. 2a,b and 21a–c). Hence, based on parsimony, and based on the fitted models, it is much more likely that indeed N grew approximately exponentially over time, with approximately constant speciation and extinction rates.

S.5 Fitting homogenous-rate models to heterogenous-rate trees

As described in the main article, over geological time global microbial diversification appears to be well-described by simple branching models with homogenous (i.e. clade-independent) speciation and extinction rates (also known as "clock-like" models). At some sufficiently fine taxonomic scale, however, speciation and extinction rates may be clade-dependent, since speciation and extinction are coupled to ecological dynamics that vary across clades. The question thus arises how homogenous speciation/extinction rates, fitted in the aforementioned models, should be interpreted. To investigate this question, we simulated trees in which speciation and extinction rates were themselves evolving (and thus clade-dependent) traits, and then fitted models with homogenous speciation/extinction rates to the corresponding LTTs. In the simulated trees, the speciation rate λ and extinction rate μ were modeled as independent Brownian motions evolving along the tree edges and constrained in a finite interval via reflection. Thus, at any moment in time the tree-wide absolute speciation (or extinction) rate was the sum of speciation (or extinction) rates of all extant tips.

Simulations were performed using the function `generate_tree_with_evolving_rates` in the R package `castor`[17]. Each simulated tree initially contained 10,000 tips, but was subsequently rarefied at some fraction ρ that was chosen randomly and uniformly between 0.1 and 1 to reflect incomplete sampling in our real trees. For each simulated tree, the maximum allowed speciation rate, $\lambda_{\max}$, was chosen randomly and

uniformly between 0 and 1 speciations per lineage per million years (S/LMa). The maximum allowed extinction rate, $\mu_{\max}$, was chosen randomly and uniformly between 0 and $\lambda_{\max}$. The minimum allowed speciation and extinction rates, $\lambda_{\min}$ and $\mu_{\min}$, were set to 0. The diffusivity of the Brownian motion model for λ, denoted D_λ, was roughly chosen such that within a given expected number of speciations along a lineage ("rate memory", denoted M), λ would drift to a standard deviation equal to $\frac{1}{2}(\lambda_{\max} + \lambda_{\min})$, that is:

$$D_\lambda = \frac{1}{2M}\left(\lambda_{\min} + \lambda_{\max}\right) \cdot \left[\frac{1}{2}\left(\lambda_{\max} + \lambda_{\min}\right)\right]^2. \tag{30}$$

The diffusivity D_μ was chosen in a similar way:

$$D_\mu = \frac{1}{2M}\left(\lambda_{\min} + \lambda_{\max}\right) \cdot \left[\frac{1}{2}\left(\mu_{\max} + \mu_{\min}\right)\right]^2. \tag{31}$$

Note that M can be interpreted as the approximate expected phylogenetic depth (number of consecutive branching points) at which λ and μ are conserved. For example, for the parameters used in this study, a rate memory of $M = 100$ corresponds to a diffusivity $D_\lambda = 0.00125\ (\mathrm{S/LMa})^2/\mathrm{Ma}$ and a phylogenetic conservatism of ∼200 Ma for λ.

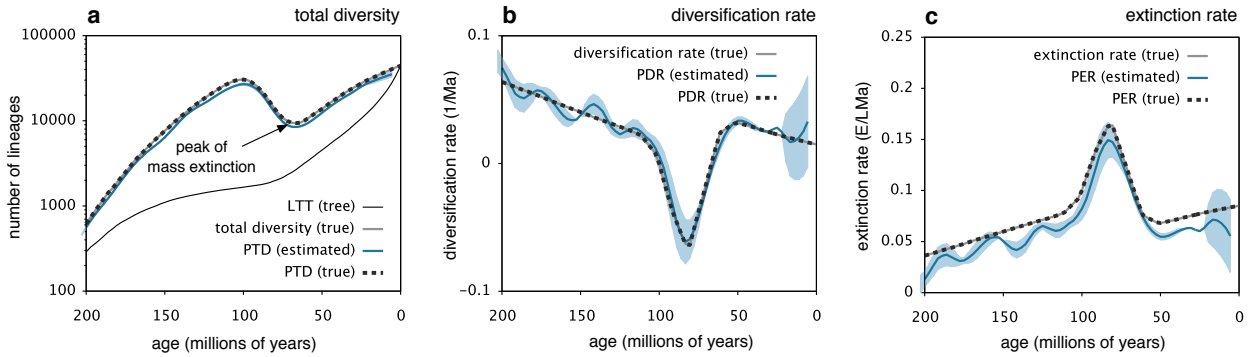
We investigated the suitability of homogenous rate models for representing speciation/extinction dynamics in these simulated trees, as well as the relationship between the fitted homogenous λ, μ and the probability distributions of the (clade-specific) λ, μ across tips of the simulated trees. During model fitting, the sampling fraction ρ applied to each simulated tree was assumed to be known. We repeated our investigation for various rate memories M, each time using 100 simulated trees.

In all cases, homogenous-rate models with constant λ and μ fitted the simulated LTTs well (mean relative deviations below 10%). This suggests that timetrees, in which speciation and extinction rates are evolving heritable traits and thus not clock-like, can still be described by models with homogenous rates.

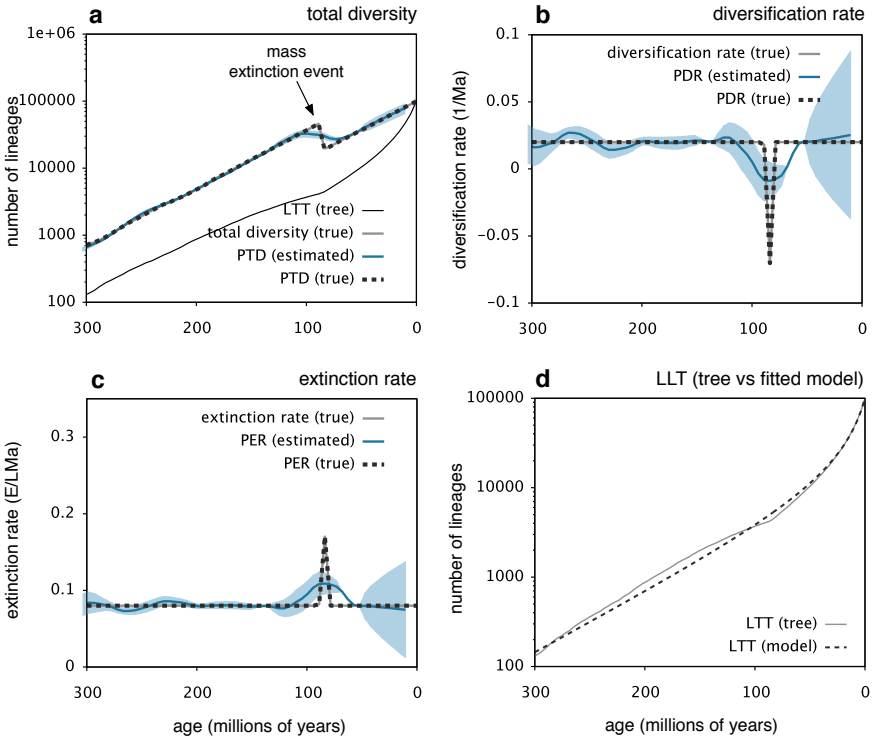
As shown in Supplementary Fig. 20, the λ and μ estimated from homogenous-rate models approximately correspond to the average λ and μ across extant tips in the tree. For λ, this approximation is quite accurate across all of our tested scenarios (mean $R^2 > 0.95$, Figs. 20a–c). For μ, the accuracy of this approximation can vary substantially, depending on the rate of evolution of λ and μ (i.e., their phylogenetic conservatism). This observation is consistent with previous findings that heterogenous diversification rates lead to inflated estimates of extinction rates when the latter are obtained from homogenous-rate models [18]. Our results show that the intensity of this bias is related to the rate of evolution of the traits determining speciation and extinction rates; the bias becomes stronger when λ and μ exhibit strong phylogenetic conservatism. This relationship can be explained by the fact that our formulas have been derived under the assumption that the probabilities of two sister-lineages going extinct are independent ($\mathbb{M}_{02}(t,T) = E(t,T)^2$ in Eq. 6). This assumption is only true if λ and μ are homogenous, or if λ and μ evolve sufficiently fast. That said, even for very strongly conserved λ and μ (rate memory ~ 100; Fig. 20f), the relative error is in the order of 80–100%, that is, even under such a scenario our methods are suitable for estimating average extinction rates to order of magnitude accuracy. In conclusion, for any given time point, rates estimated here should be interpreted as approximate current averages over the entire tree, but it is unknown whether rates were homogenous (i.e., equal to the average) or heterogenous (i.e., exhibiting a variance around the average) at that time point.

S.6 Implications for reconstructing diversification from phylogenies

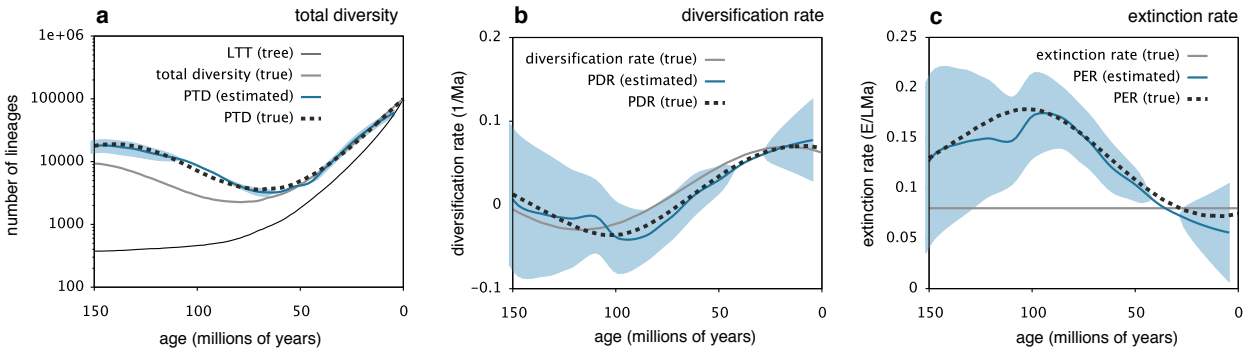
Debate exists over the information that can be extracted from molecular phylogenies in the absence of fossil data, with some authors even suggesting that extinctions cannot possibly be reconstructed from phylogenies alone[8,18–20]. Indeed, in principle similar LTTs (or branch length distributions) can be generated under different combinations of speciation and extinction rates varying over time in specific ways. In the most extreme case, a tree generated by a speciation-extinction process could in principle also be the result of a pure speciation process, with a particular idiosyncratically varying speciation rate over time generating the same LTT. This ambiguity is reflected in the PDR — a curve with similar information content as the LTT, representing an inseparable combination of the diversification rate and the relative rate of change of the speciation rate (Eq. 1). The problem is further amplified in some analyses that use simple (single-valued) summary statistics, such as the "γ-statistic", to describe the overall shape of the LTT and then attempt to associate different values of the statistic with different diversification scenarios[6,19,21]. Importantly, molecular phylogenies completely lack information on the diversification dynamics of clades that are now entirely extinct. Hence, any reconstruction of past bacterial diversification assumes that the speciation/extinction rates within extinct clades were similar to (or distributed similarly to, if non-homogenous) those of extant clades. Despite the limitations of molecular phylogenies and reservations by some authors[8,18], extinction does leave characteristic traces in phylogenies that tend to be markedly different from typical scenarios lacking extinction, as demonstrated here (e.g., Fig. 1) and elsewhere[20]. Molecular phylogenies thus contain information on past extinctions, albeit in a convoluted format, and this information can be extracted using methods such as ours.



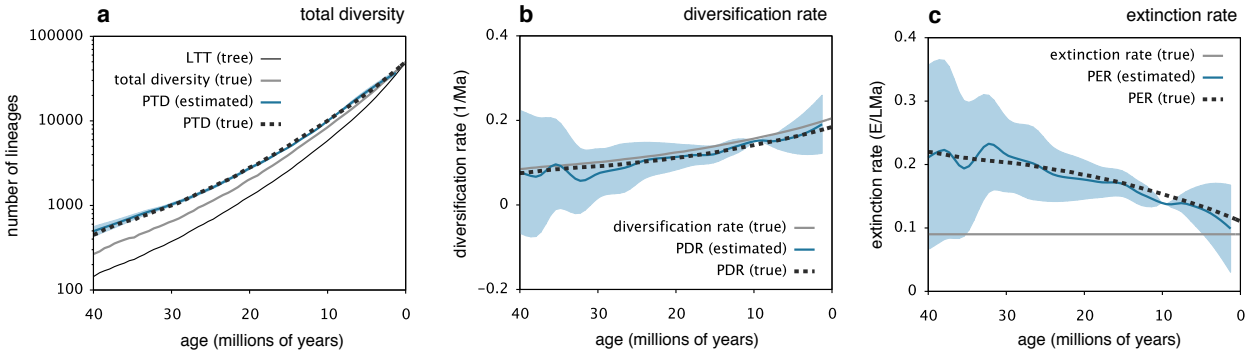
Supplementary Figure 1: **Demonstration of non-parametric estimation (mass extinction event).** Simulation and non-parametric analysis of a tree with constant speciation rate and variable extinction rate (including diversity dependence and a mass extinction event about 80 Ma ago). (a) Lineages through time (LTT) in the tree (black continuous curve), true total diversity (grey continuous curve), true pulled total diversity (PTD, dashed curve) and estimated PTD (blue continuous curve). The footprint of the mass extinction event is visible in the tree's LTT. (b) Diversification rate (grey continuous curve), true pulled diversification rate (PDR, dashed curve) and estimated PDR (blue continuous curve, Eq. 25). (c) True extinction rate (grey continuous curve), true pulled extinction rate (PER, dashed curve) and estimated PER (blue continuous curve). Estimated quantities in A–C were noise filtered similarly to the bacterial trees (Fig. 2); blue shades indicate standard errors of the noise filter. All "pulled" variables (PTD, PDR, PER) are equivalent to their non-pulled variants (total diversity, diversification rate, extinction rate) because the speciation rate is constant over time; consequently, curves showing true pulled variables (dashed lines) overlap completely with curves showing non-pulled variables (grey continuous lines).



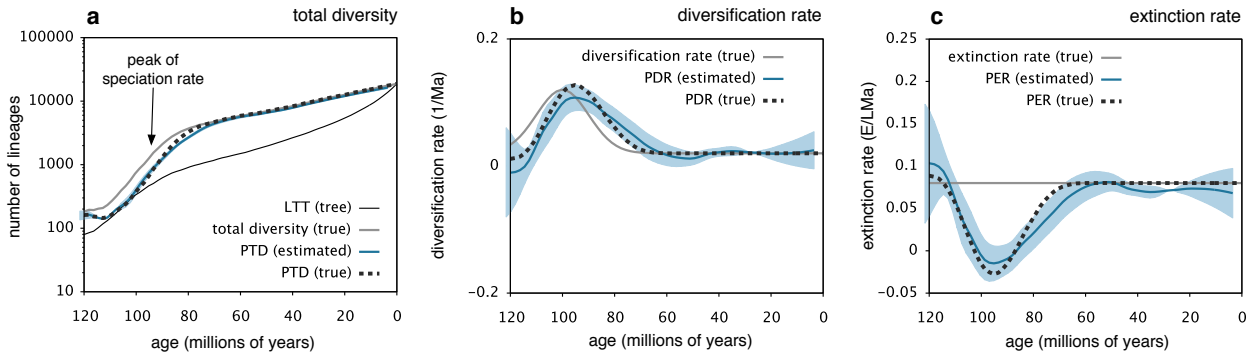
Supplementary Figure 2: **Demonstration of non-parametric estimation (sharp mass extinction event).** Simulation and non-parametric analysis of a tree with constant speciation rate and an extinction rate subject to a sharp temporary increase about 400 Ma ago (representing a sharp mass extinction event). (a) Lineages through time (LTT) in the tree (black continuous curve), true total diversity (grey continuous curve), true pulled total diversity (PTD, dashed curve) and estimated PTD (blue continuous curve). (b) Diversification rate (grey continuous curve), true pulled diversification rate (PDR, dashed curve) and estimated PDR (blue continuous curve, Eq. 25). (c) True extinction rate (grey continuous curve), true pulled extinction rate (PER, dashed curve) and estimated PER (blue continuous curve). (d) LTT in the tree (continuous curve), compared to the LTT predicted by a fitted constant-rates model (dashed curve). The deviation of the model's prediction from the tree's LTT is most clear as an inflection point at age $\sim$400 Ma. Estimated quantities in (a–c) were noise filtered similarly to the bacterial trees (Fig. 2); blue shades indicate standard errors of the noise filter. The sharp spike in the extinction rate introduces a sharp spike in the PDR (b) and PER (c), which, however, is damped by the noise filter during estimation. All "pulled" variables (PTD, PDR, PER) are equivalent to their non-pulled variants (total diversity, diversification rate, extinction rate) because the speciation rate is constant over time; consequently, curves showing true pulled variables (dashed lines) overlap completely with curves showing non-pulled variables (grey continuous lines).



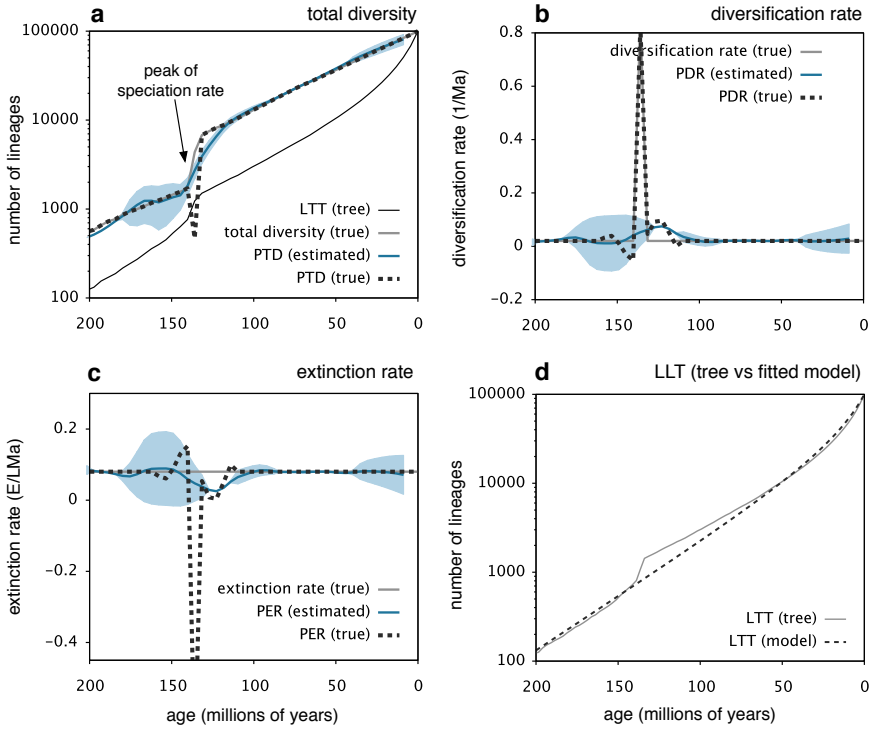
Supplementary Figure 3: **Demonstration of non-parametric estimation (oscillating speciation rate).** (a) Lineages through time for a simulated tree with strongly oscillating speciation rate ($\pm 50\%$ around the mean) and variable extinction rate (including diversity dependence and a mass extinction event about 60 Ma in the past). Dashed curve: Lineages through time (LTT) in the tree. Grey continuous curve: True total diversity. Blue continuous curve: Pulled total diversity (PTD), estimated from the LTT. The fluctuations in the speciation rate lead to a deviation of the PTD from the total diversity. Nevertheless, the mass extinction event is clearly reflected in the PTD. (b) Diversification rate (thin curve), true pulled diversification rate (PDR, dashed curve) and estimated PDR (blue curve, Eq. 25), during the same simulation as in (a). (c) True extinction rate (grey curve) and pulled extinction rate (PER, blue curve) estimated from the LTT (Eq. 22). The fluctuations in the speciation rate clearly distort the pulled extinction rate. All estimated quantities were noise filtered similarly to the bacterial trees (Fig. 2); blue shades indicate standard errors of the noise filter.



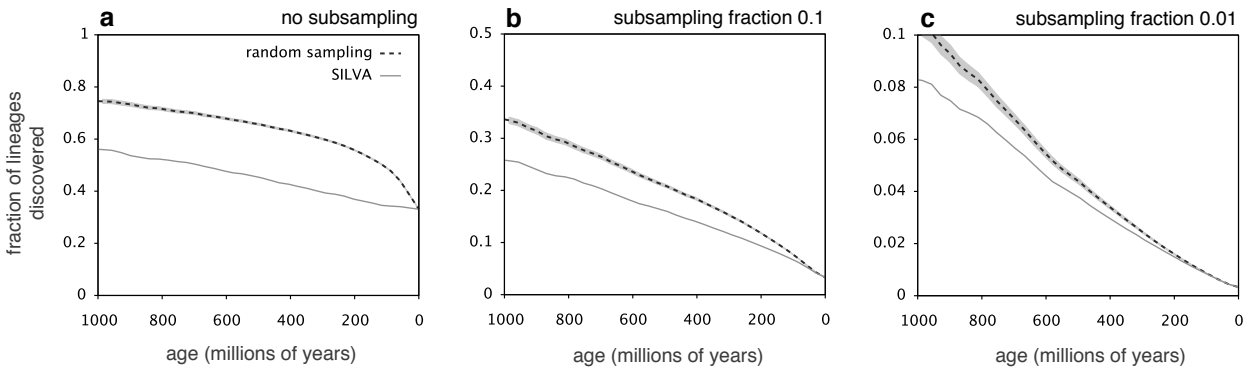
Supplementary Figure 4: **Demonstration of non-parametric estimation (diversity-dependent speciation rate).** (a) Lineages through time for a simulated tree with density-dependent speciation rate ($\lambda \propto N^{0.1}$) and constant extinction rate. Dashed curve: Lineages through time (LTT) in the tree. Grey continuous curve: True total diversity. Blue continuous curve: Pulled total diversity (PTD), estimated from the LTT. (b) Diversification rate (thin curve), true pulled diversification rate (PDR, dashed curve) and estimated PDR (blue curve, Eq. 25), during the same simulation as in (a). (c) True extinction rate (grey curve) and pulled extinction rate (PER, blue curve) estimated from the LTT (Eq. 22). The density-dependence of the speciation rate introduces a trend in the pulled extinction rate. All estimated quantities were noise filtered similarly to the bacterial trees (Fig. 2); blue shades indicate standard errors of the noise filter.



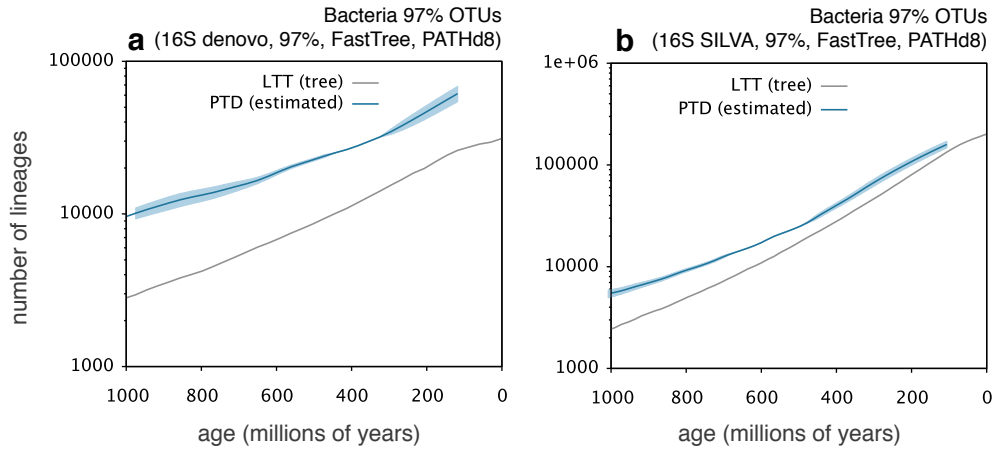
Supplementary Figure 5: **Demonstration of non-parametric estimation (speciation spike).** Simulation and non-parametric analysis of a tree with a temporarily increased speciation rate about 90 million years ago and a constant extinction rate. (a) Lineages through time (LTT) in the tree (black continuous curve), true total diversity (grey continuous curve), true pulled total diversity (PTD, dashed curve) and estimated PTD (blue continuous curve). (b) Diversification rate (grey continuous curve), true pulled diversification rate (PDR, dashed curve) and estimated PDR (blue continuous curve, Eq. 25), during the same simulation as in (a). (c) True extinction rate (grey continuous curve), true pulled extinction rate (PER, dashed curve) and estimated PER (blue continuous curve). The temporarily increased speciation rate introduces a similarly strong distortion in the PER. All estimated quantities were noise filtered similarly to the bacterial trees (Fig. 2); blue shades indicate standard errors of the noise filter.



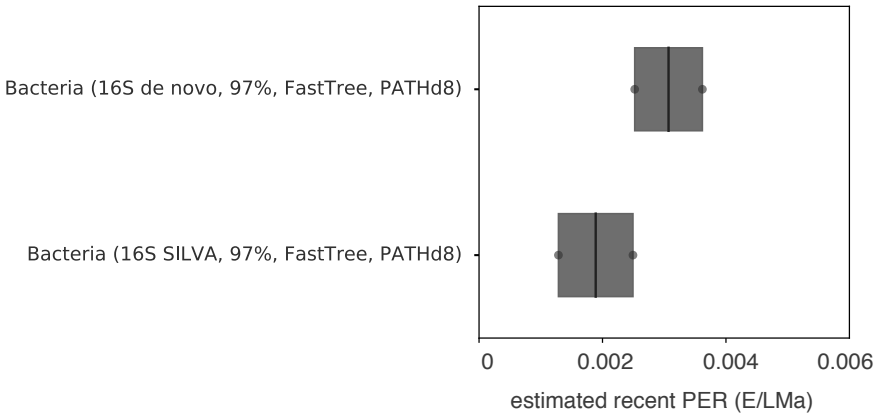
Supplementary Figure 6: **Demonstration of non-parametric estimation (brief speciation spike).** Simulation and non-parametric analysis of a tree with a briefly but strongly increased speciation rate about 140 million years ago and a constant extinction rate. (a) Lineages through time (LTT) in the tree (black continuous curve), true total diversity (grey continuous curve), true pulled total diversity (PTD, dashed curve) and estimated PTD (blue continuous curve). (b) Diversification rate (grey continuous curve), true pulled diversification rate (PDR, dashed curve) and estimated PDR (blue continuous curve, Eq. 25), during the same simulation as in (a). (c) True extinction rate (grey continuous curve), true pulled extinction rate (PER, dashed curve) and estimated PER (blue continuous curve). (d) LTT in the tree (continuous curve), compared to the LTT predicted by a fitted constant-rates model (dashed curve). The deviation between the tree's LTT and the model prediction is clearly visible. Estimated quantities in (a–c) were noise filtered similarly to the bacterial trees (Fig. 2); blue shades indicate standard errors of the noise filter. The brief spike in the speciation rate introduces a sharp spike in the PDR (b) and PER (c), which, however, is damped by the noise filter during estimation. Note that curves showing true pulled variables (dashed lines) overlap largely with curves showing non-pulled variables (grey continuous lines).



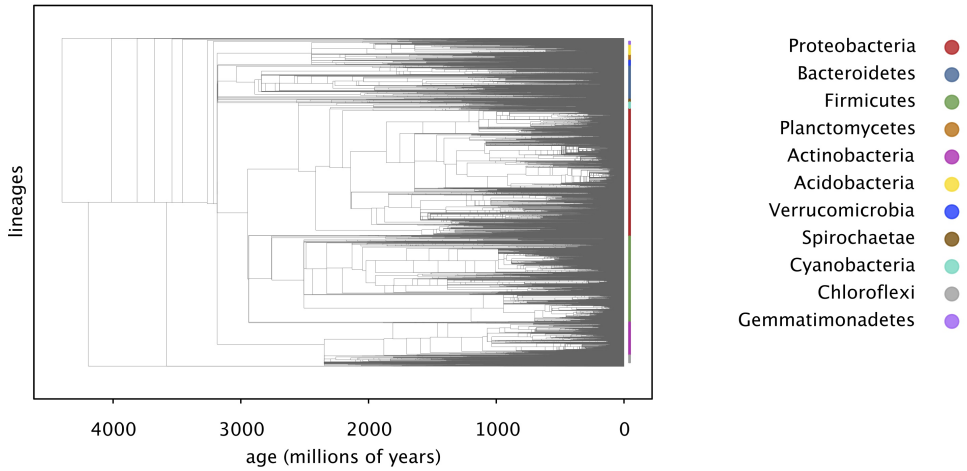
Supplementary Figure 7: **Fractions of discovered lineages over time (SILVA vs random OTU sampling).** (a) Fraction of bacterial lineages in the *de novo* dataset (162,371 de-novo OTUs at 99% 16S rRNA similarity), previously discovered (i.e. matched to SILVA at 99% similarity), as a function of lineage age (continuous line). The dashed line shows the expectation under the null model of random non-phylogenetically biased OTU discovery and the shading indicates the corresponding standard deviation. (b) Similar to a, but after subsampling SILVA down to a fraction of 10%. (c) Similar to a, but after subsampling SILVA down to a fraction of 1%. The deviation from the null model is due to a non-random (phylogenetically biased) representation of clades in SILVA.



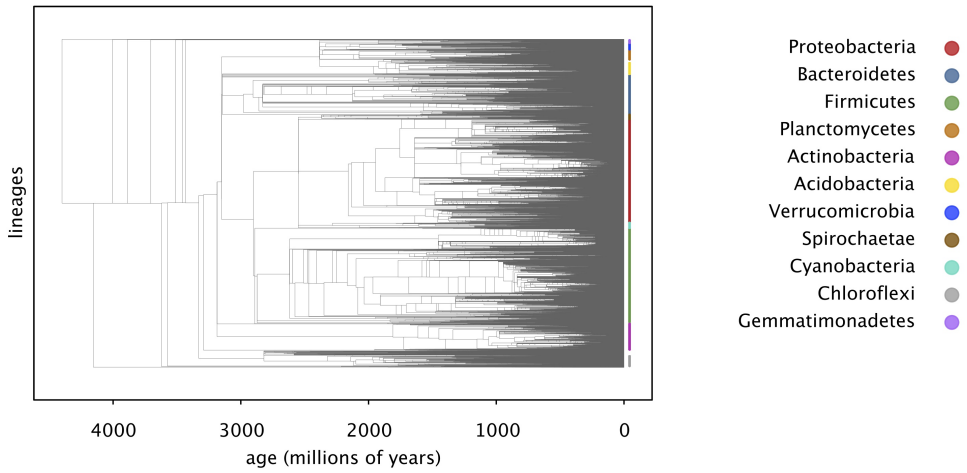
Supplementary Figure 8: **Bacterial diversities over time (based on 97%-OTUs).** Lineages through time (LTT, grey curves), compared to estimated pulled total diversities (PTD, blue curves) of bacteria, based on timetrees of 16S rRNA OTUs delineated at 97% similarity. (a) Based on partial-length (V4) *de novo* 97%-OTUs. (b) Based on full-length 97%-OTUs from SILVA. Summaries of timetree construction methods are indicated in brackets.



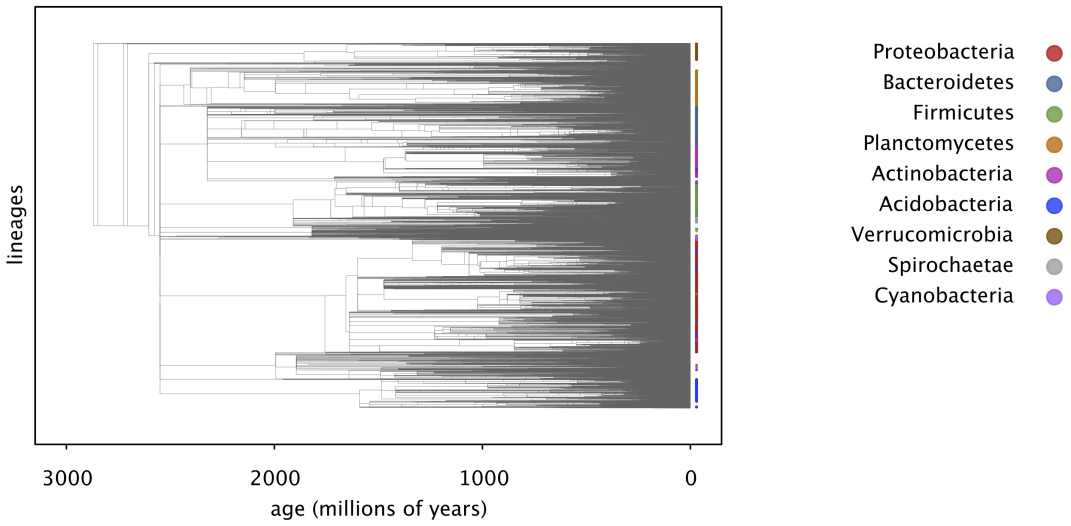
Supplementary Figure 9: **Recent pulled extinction rates (bacterial 97%-OTUs).** Non-parametrically estimated recent pulled extinction rates of bacteria, based on timetrees of 16S rRNA OTUs delineated at 97% similarity. Each box corresponds to a different tree, and comprises results obtained by assuming various numbers of total extant OTUs (Supplementary Table 6). Summaries of timetree construction methods are indicated in brackets.



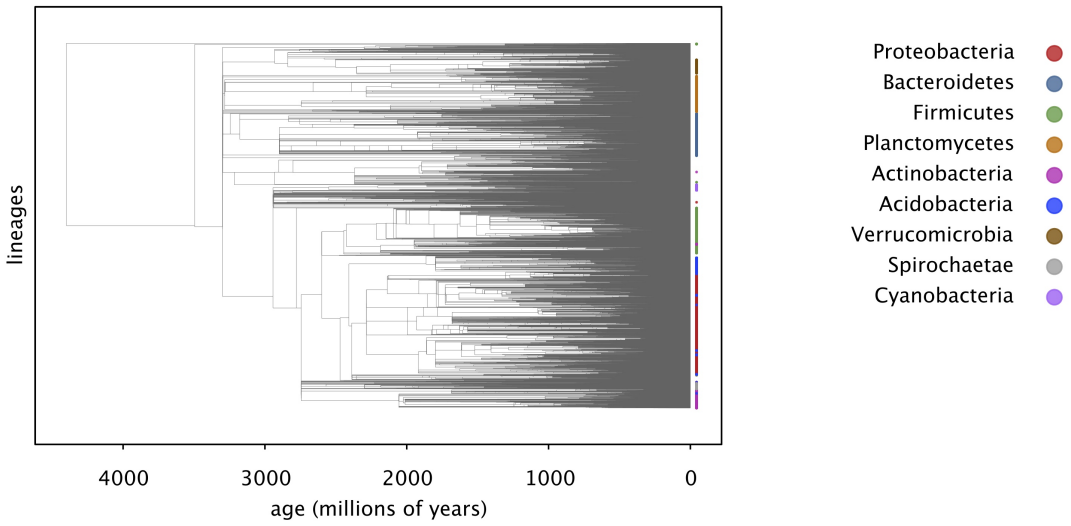
Supplementary Figure 10: Bacterial timetree (16S SILVA, FastTree, PATHd8), used for the diversification analysis. Major phyla (based on SILVA 128)[22] are indicated as colored segments.



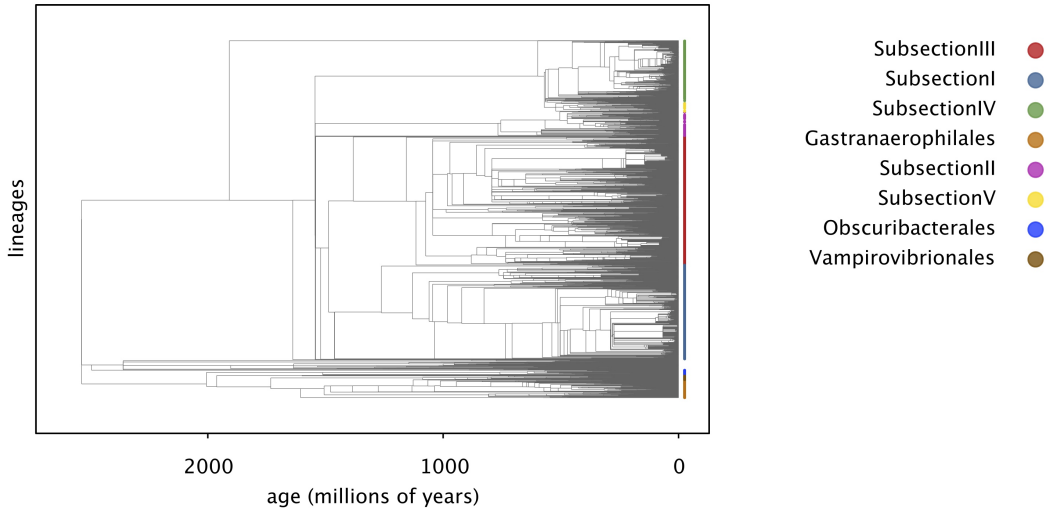
Supplementary Figure 11: Bacterial timetree (16S SILVA, 97%, FastTree, PATHd8), used for the diversification analysis. Major phyla (based on SILVA 128)[22] are indicated as colored segments.



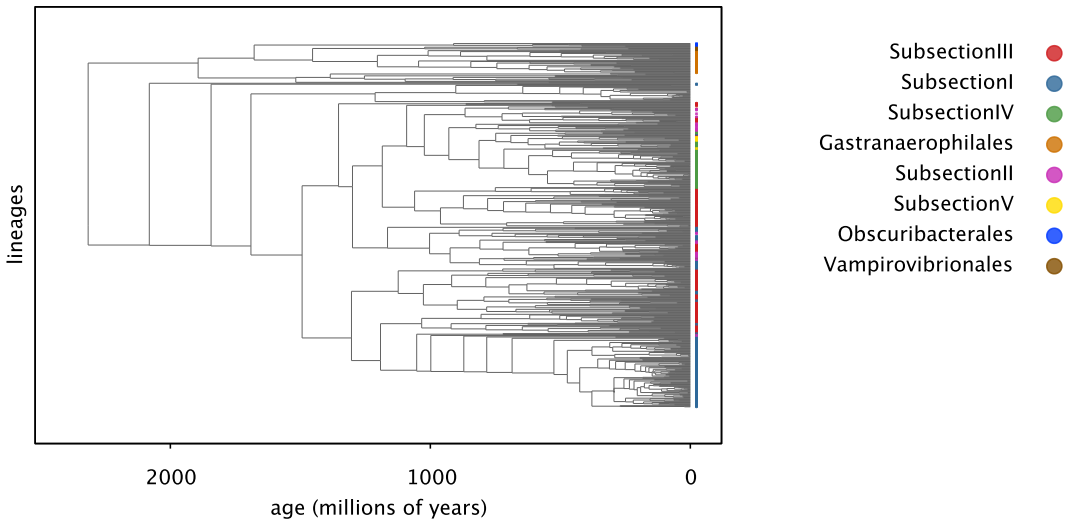
Supplementary Figure 12: Bacterial timetree (16S V4 de novo, FastTree, PATHd8), used for the diversification analysis. Major phyla (based on SILVA 128)[22] are indicated as colored segments. Note that some tips belong to unidentified phyla, and are thus not highlighted in color. Also note that the phylogeny is not absolutely congruent with taxonomic assignments.



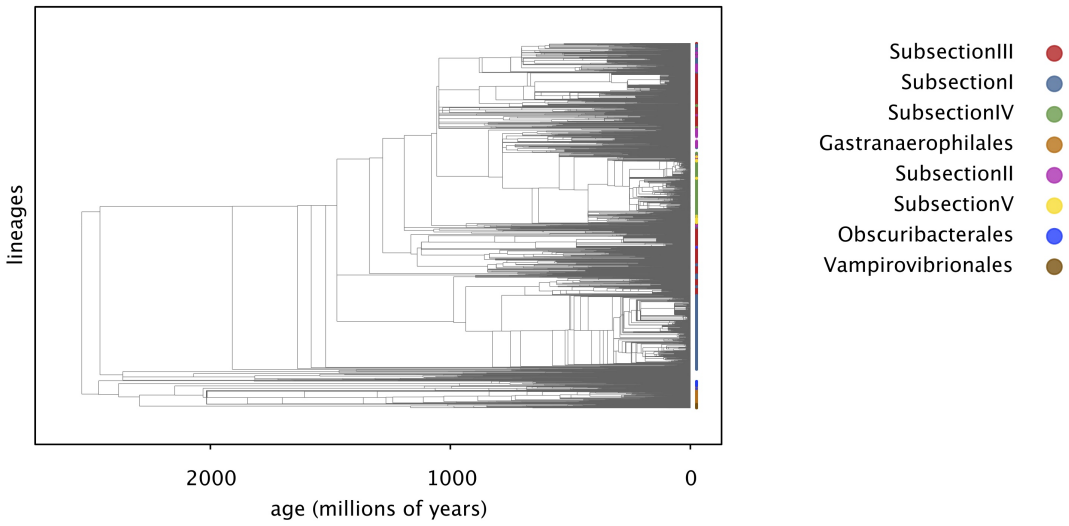
Supplementary Figure 13: Bacterial timetree (16S de novo, 97%, FastTree, PATHd8), used for the diversification analysis. Major phyla (based on SILVA 128)[22] are indicated as colored segments. Note that some tips belong to unidentified phyla, and are thus not highlighted in color. Also note that the tree is not absolutely congruent with taxonomic assignments.



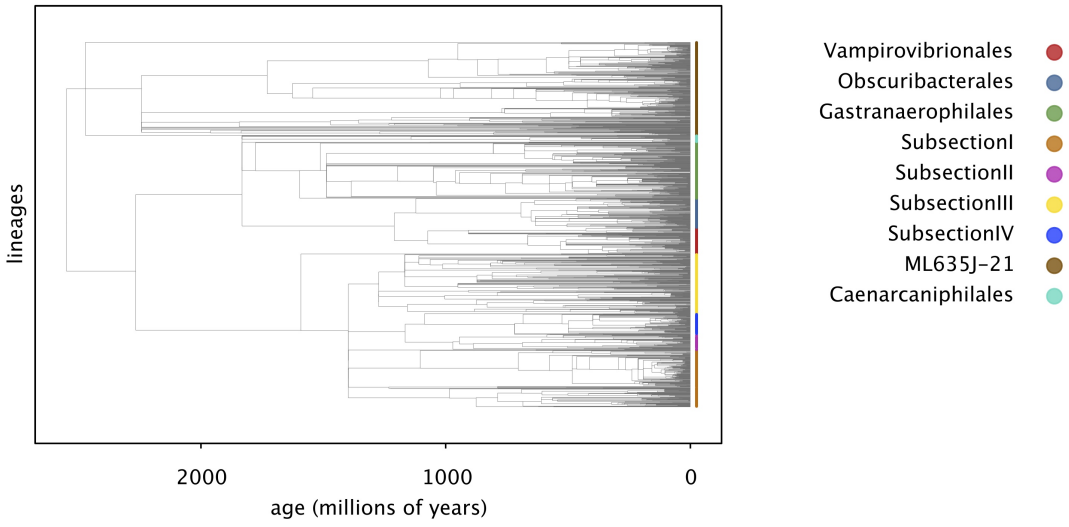
Supplementary Figure 14: Cyanobacterial timetree (16S SILVA, FastTree, BEAST+PATHd8), used for the diversification analysis. Major classes or orders (based on SILVA 128)[22] are indicated as colored segments. Note that some tips belong to unidentified classes or orders, and are thus not highlighted in color.



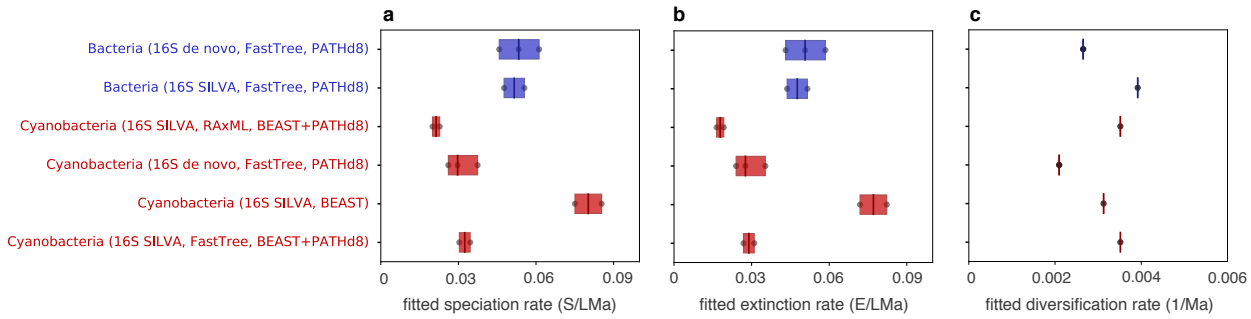
Supplementary Figure 15: Cyanobacterial timetree (16S SILVA, BEAST), used for the diversification analysis. Major classes or orders (based on SILVA 128)[22] are indicated as colored segments. Note that some tips belong to unidentified classes or orders, and are thus not highlighted in color. Also note that the tree is not absolutely congruent with taxonomic assignments.



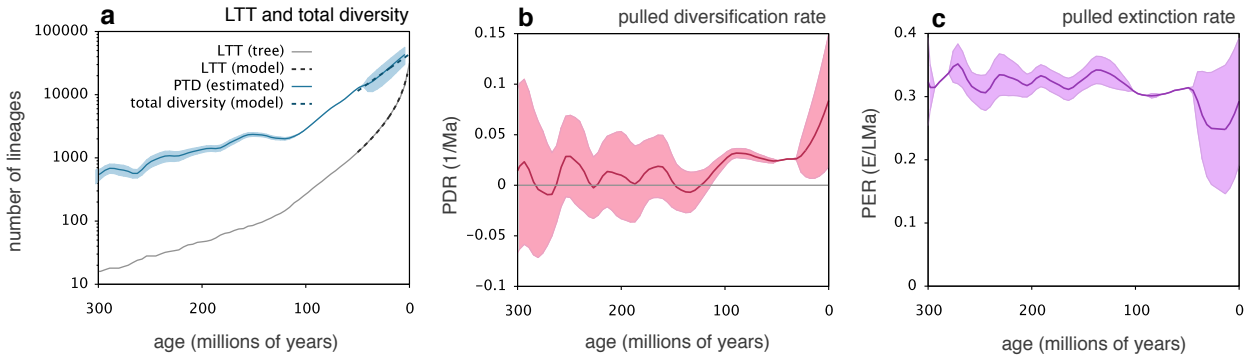
Supplementary Figure 16: Cyanobacterial timetree (16S SILVA, RAXML, BEAST+PATHd8), used for the diversification analysis. Major classes or orders (based on SILVA 128)[22] are indicated as colored segments. Note that some tips belong to unidentified classes or orders, and are thus not highlighted in color. Also note that the tree is not absolutely congruent with taxonomic assignments.



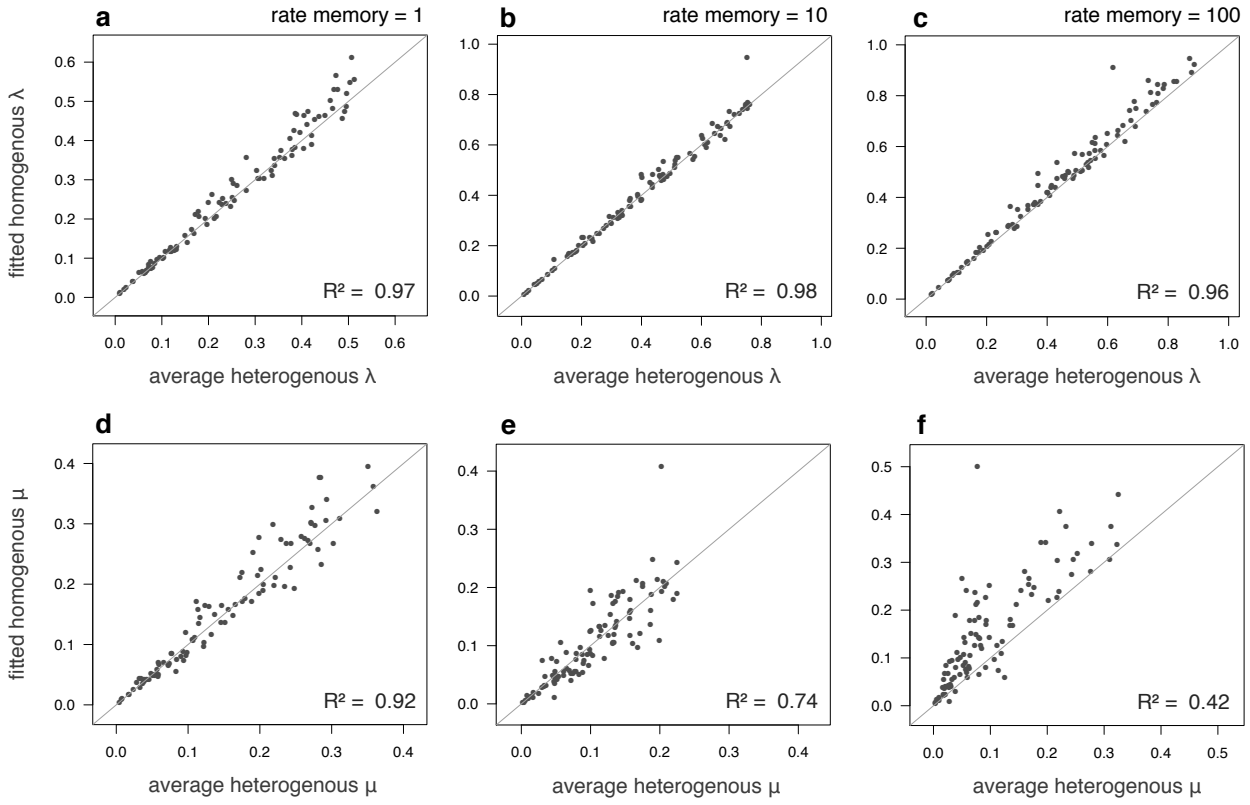
Supplementary Figure 17: Cyanobacterial timetree (16S de novo, FastTree, PATHd8), used for the diversification analysis. Major classes or orders (based on SILVA 128)[22] are indicated as colored segments.



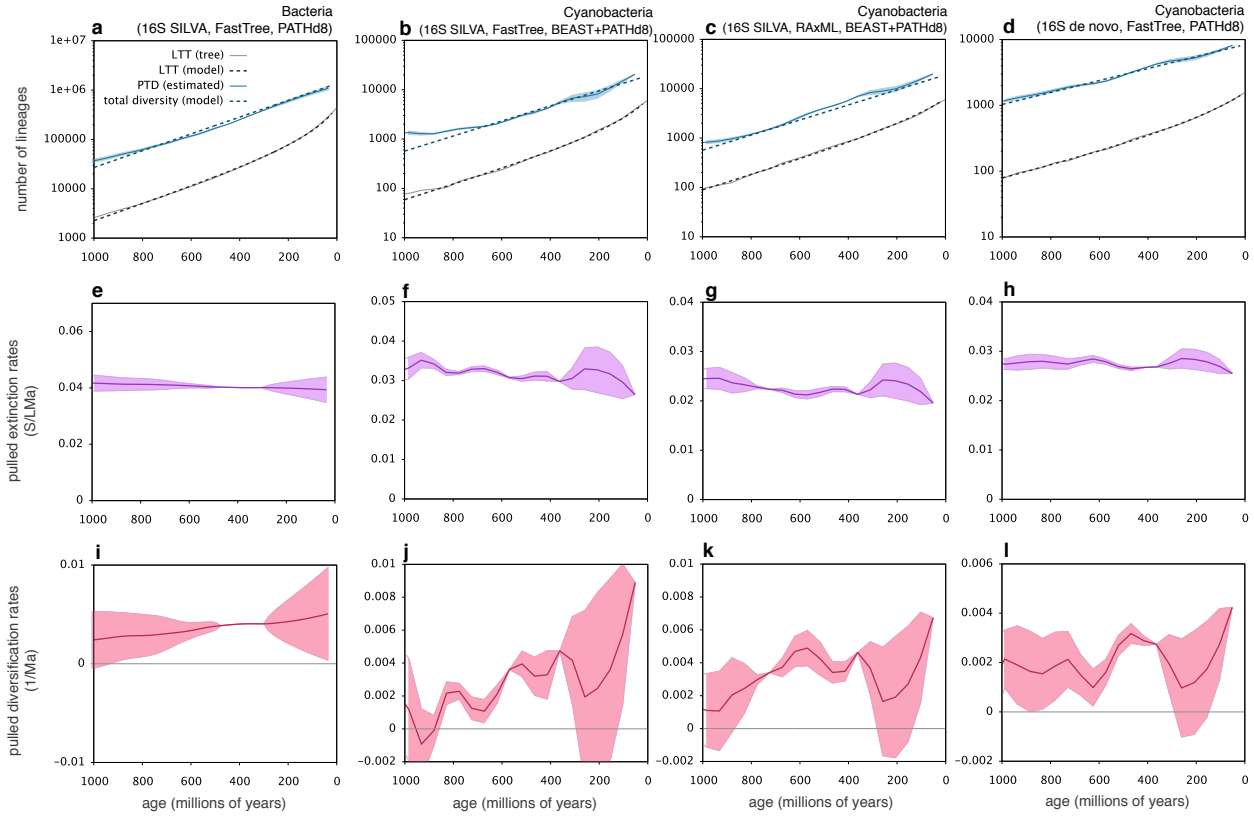
Supplementary Figure 18: **Estimated long-term speciation, extinction and diversification rates.** Speciation rates (a), extinction rates (b) and diversification rates (c), estimated for various taxa and using various dated timetrees (one box per timetree). Estimates are obtained by fitting cladogenic models over the past 1 billion years (for estimates over shorter time intervals see Fig. 3 in the main article). Each box comprises results obtained by assuming various numbers of total extant OTUs (Supplementary Tables 2, 4 and 3). Tree labels and boxes are colored by taxon. Summaries of timetree sources or construction methods are indicated in brackets (see Methods for details).



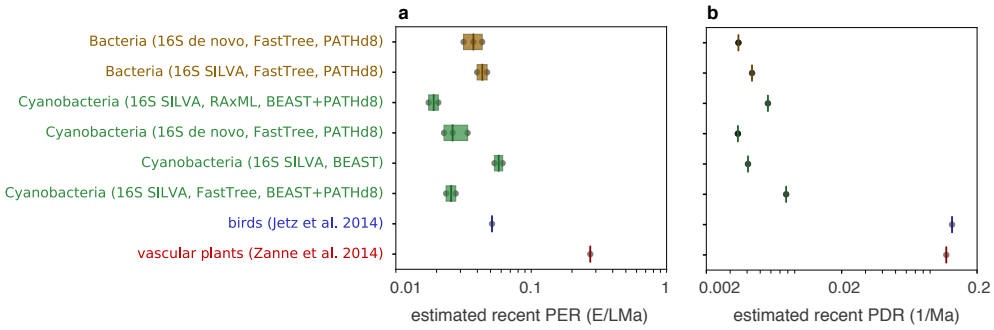
Supplementary Figure 19: **Diversification of vascular plants over time.** (a): Lineages through time for vascular plants (grey continuous line), compared to a speciation-extinction model fitted to a recent age interval (grey dashed lines). The blue continuous curve shows non-parametrically estimated pulled total diversities (PTDs), the blue dashed curve shows total diversities predicted by the fitted model. (b) Pulled extinction rates over time, estimated non-parametrically. (c): Pulled diversification rates, estimated non-parametrically. The tree was taken from Zanne *et al.* [23].



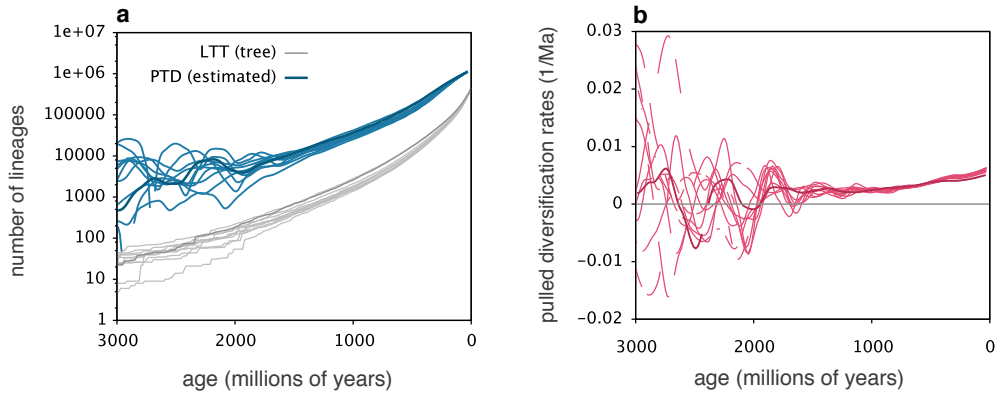
Supplementary Figure 20: **Fitting homogenous-rate models to heterogenous-rate trees.** (a–c) Average speciation rates across extant tips of simulated trees with evolving λ (horizontal axes), compared to fitted homogenous speciation rates (vertical axis). One point pet simulated tree. Rate memories of speciation rates were 1 in (a), 10 in (b) and 100 in (c); recall that a lower rate memory corresponds to a faster evolving λ and μ. Diagonals are shown for reference. Fractions of variance in the vertical axis explained by the horizontal axis (R^2) are written in each figure. (d–f) Analogous to (a–c), but showing extinction rates. Methodological details provided in Supplement S.5.



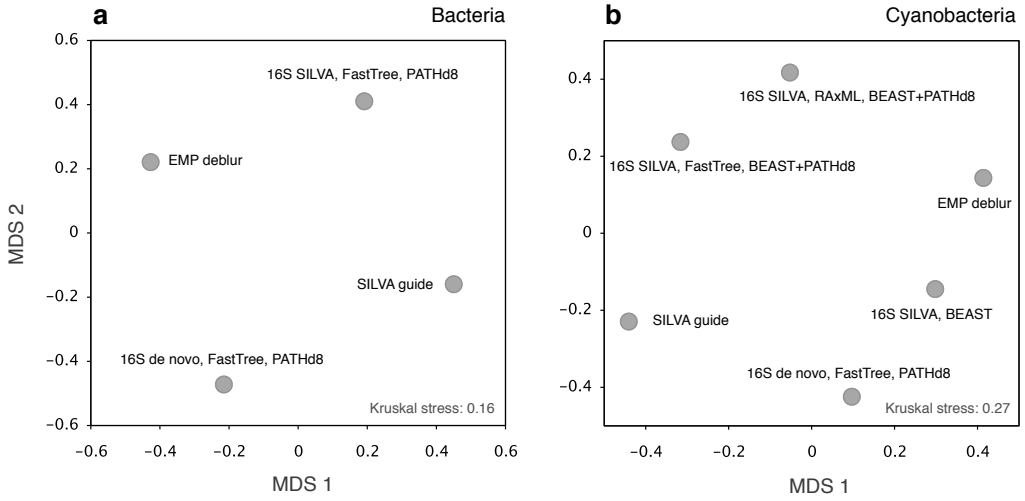
Supplementary Figure 21: **Bacterial and cyanobacterial diversification over time (alternative trees).** (a–d): Lineages through time (LTT) of a bacterial timetree (a) and three alternative cyanobacterial timetrees (b–d), compared to predictions by speciation-extinction models fitted over the last 1 billion years (grey dashed curves). Blue continuous curves show non-parametrically estimated pulled total diversities (PTDs), blue dashed curves show total diversities predicted by the fitted models. Note that each tree only comprises a subset of extant species, and thus the true extant diversity (right-most point on blue curves) is only an estimate (overview in Table 2). (e–h) Pulled extinction rate (PER) over time, estimated non-parametrically from the trees used in (a–d) (Supplement S.1.3). (i–l): Pulled diversification rate (PDR), estimated non-parametrically from the same trees as in (a–d). In all figures, shades indicate standard errors of noise-filtered estimates. Summaries of tree construction methods are indicated in brackets (see Methods for details).



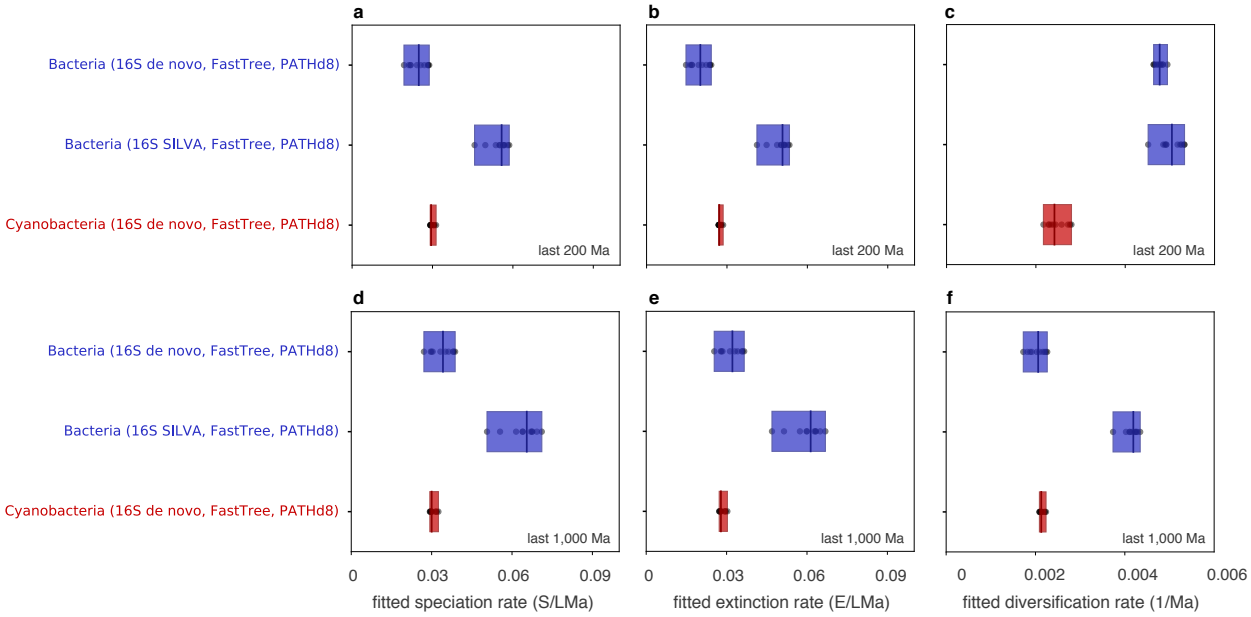
Supplementary Figure 22: **Estimated recent pulled extinction and diversification rates.** Recent pulled extinction rates (a) and recent pulled diversification rates (b), estimated non-parametrically for various taxa and using various dated timetrees (one box per timetree). Each box corresponds to a different tree, and comprises results obtained by assuming various numbers of extant bacterial or cyanobacterial OTUs (Supplementary Tables 2, 4 and 3). Tree labels and boxes are colored by taxon. Summaries of timetree sources or construction methods are indicated in brackets (see Methods for details).



Supplementary Figure 23: **Bacterial diversities over time (sensitivity to dating).** (a) Lineages through time (LTT, grey curves), compared to estimated pulled total diversities (PTD, blue curves) of bacteria, based on timetrees constructed from 16S sequences in SILVA (timetree "16S SILVA, FastTree, PATHd8" and random variants). Individual curves correspond to slightly different timetrees, dated using randomly varied dating constraints for purposes of sensitivity analysis (see Methods for details). (b) Estimated pulled diversification rates (PDR) for the same trees as in (a). In both figures, darker curves correspond to the point-estimate timetree ("16S SILVA, FastTree, PATHd8") discussed in the main article. Observe that the variability between tree variants as well as spurious fluctuations, and thus the uncertainty in the estimated PTD and PDR, increases drastically for ages older than 1.5 billion years.



Supplementary Figure 24: **Comparison of tree topologies.** Multidimensional scaling plot showing pairwise distances between bacterial (a) or cyanobacterial (b) timetrees investigated in the main article (one point per tree). Points closer to each other indicate a greater similarity. Tree distances are based on the Robinson-Foulds metric[24], which measures the difference in tree topology when restricted to tips common to both trees being compared. Tree summaries are indicated next to each point (see Methods for details). "EMP deblur" is taken from[10], and "SILVA guide" refers to the SILVA SSU guide tree (release 128)[22]. For a list of pairwise tree distances, see Supplementary Table 7.



Supplementary Figure 25: **Sensitivity of rate estimates to dating constraints.** Speciation rates (a,d), extinction rates (b,e) and diversification rates (c,f), estimated for Bacteria or Cyanobacteria, using various dated trees (one box per tree). Estimates are obtained by fitting cladogenic models over the last 200 million years (top row) or the last 1 billion years (bottom row). Each box comprises results obtained by varying dating anchors within their uncertainty intervals (Supplementary Table 5, details in Methods). Tree labels and boxes are colored by taxon. Summaries of tree sources or construction methods are indicated in brackets. Only timetrees dated using PATHd8 and based on primary constraints (Table 5) are shown.

Supplementary Table 1: **Overview of timetrees.** Overview of timetrees analyzed in the main article, including source or construction method, tree size (number of tips), the age interval considered for estimating recent speciation/extinction rates (via model fitting, Supplement S.1.2), fitted recent speciation/extinction rates (λ and μ) and mean relative deviation (MRD) of models fitted over 1 billion years. Each tree's sampling fraction (ρ) was set to its size divided by the corresponding (independently estimated) total number of extant OTUs, as listed in Table 2.

tree	**size (tips)**	**ages (Ma)**	λ **(S/LMa)**	μ **(E/LMa)**	**MRD**
Bacteria (16S SILVA, FastTree, PATHd8)	448,112	20–200	0.0440	0.0393	0.030
Bacteria (16S de novo, FastTree, PATHd8)	162,371	20–200	0.0350	0.0300	0.025
Cyanobacteria (16S SILVA, BEAST)	586	20–200	0.0651	0.0594	0.024
Cyanobacteria (16S de novo, FastTree, PATHd8)	1,579	20–200	0.0294	0.0271	0.020
Cyanobacteria (16S SILVA, FastTree, BEAST+PATHd8)	6,308	20–200	0.0340	0.0309	0.046
Cyanobacteria (16S SILVA, RAxML, BEAST+PATHd8)	6,302	20–200	0.0255	0.0226	0.035
birds[25]	9,993	0–20	0.172	0.0828	–
vascular plants[23]	31,749	5–50	0.332	0.306	–

Supplementary Table 2: Column 2: Fraction of *de novo* bacterial OTUs (99% similarity in the 16S rRNA gene) that were found to be represented in the SILVA v128 database (at 99% similarity). Columns 3 & 4: Number of full-length (FL) clusters and partial-length (V4) clusters, respectively, in SILVA (16S rRNA sequences clustered at 99% similarity). Columns 5 & 6: Global number of extant full-length and partial-length OTUs, respectively, based on the fraction of *de novo* OTUs represented in SILVA (columns 3 & 4 divided by column 2). See Methods for details.

taxon	**fraction *de novo* OTUs**	**FL clusters in SILVA**	**V4 clusters in SILVA**	**estimated extant FL**	**estimated extant V4**
Bacteria	0.329	448,112	148,471	1,360,260	450,691
Cyanobacteria	0.330	6,308	2,770	19,127	8,399
Firmucutes	0.442	115,864	34,275	264,416	77,628
Bacteroidetes	0.377	45,351	16,896	120,265	44,806
Proteobacteria	0.367	174,646	47,027	476,125	128,206
Acidobacteria	0.468	12,676	5,448	27,102	11,648
Spirochaetes	0.177	3,683	2,103	20,793	11,872
Planctomycetes	0.181	8,038	5,659	44,209	31,125

Supplementary Table 3: Columns 2 & 3: Number of *de novo* OTUs and number of OTUs generated from the EMP dataset, respectively. Column 4: Overlap between *de novo* OTUs and EMP OTUs (fraction of EMP OTUs matched to *de novo* OTUs at 99% identity). Column 5: Global number of extant partial-length (V4) OTUs, estimated based on the overlap between *de novo* OTUs and EMP OTUs (column 2 divided by column 4). Column 6: Estimated global number of extant full-length (FL) OTUs, based on the previously estimated number of extant V4-OTUs and the ratio of full-length over V4 OTUs in SILVA (Table 2).

taxon	# *de novo* OTUs	# EMP OTUs	overlap *de novo* vs EMP	extant V4 OTUs	extant FL OTUs
Bacteria	165,422	333,524	0.270	612,674	1,849,358
Cyanobacteria	1,598	5,547	0.149	10,725	24,424

Supplementary Table 4: Column 2: Fraction of metagenome-assembled-genome (MAG) 16S rRNA sequences that were found to be represented in the SILVA v128 database (at 99% similarity). Column 3: Number of full-length 16S rRNA sequence clusters in SILVA (at 99% similarity) within each considered taxon. Column 4: Global number of extant full-length OTUs, estimated based on the fraction of MAGs represented in SILVA (column 3 divided by column 2). See Methods for details.

taxon	fraction MAGs	FL clusters in SILVA	estimated extant FL
Bacteria	0.283	448,112	1,583,740
Cyanobacteria	0.375	6,308	16,821

Supplementary Table 5: **Prokaryotic dating anchors.** Anchors used to date bacterial trees and the *de novo* tree. Each anchor was defined as the most recent common ancestor (MRCA) of one or more taxa. For BEAST-calibrated trees, all age intervals had a uniform prior. For PATHd8, no prior was specified. Note that some trees only contained a subset of these anchors. Also note that, because PATHd8[26] requires at least one anchor with fixed age, for PATHd8-dated trees the GOE anchor was fixed to an age of 2.55 Ga[27].

ID	MRCA	range (Ga)	description
GOE	Oxyphotobacteria, Melainabacteria	2.238–2.63	Great Oxygenation Event (GOE) and mol. clock analysis[27,28]
Chl	Chloroplasts	1.047–4.4	Rhodophyte (red algae), Bangiomorpha[29]
Ri	Rickettsiales	1.6–4.4	ancestor of mitochondrion[30]
CB	Chlorobium, Bacteroidetes	1.64–4.4	Chlorobium-specific biomarkers[31]
Chr	Chromatiaceae	1.64–4.4	purple sulfur bacteria (gammaproteobacteria) biomarkers[31]
LUCA	Archaea, Bacteria	3.5–4.4	stromatolites < LUCA < detrital zircons[32,33]

Supplementary Table 6: Fraction of bacterial *de novo* 97%-OTUs (16S rRNA sequence clusters at 97% similarity) and metagenome-assembled-genome (MAG) 16S rRNA sequence sequences that were found to be represented in the SILVA v128 database (at 97% similarity). Columns 4 & 5: Number of full-length (FL) clusters and partial-length (V4) clusters, respectively, in SILVA (16S rRNA gene sequences clustered at 97% similarity). Columns 6 & 7: Global number of extant partial-length and full-length OTUs, respectively, estimated based on the fraction of *de novo* OTUs represented in SILVA (columns 4 & 5 divided by column 2). Column 8: Global number of extant full-length OTUs, estimated based on the fraction of MAGs represented in SILVA (column 4 divided by column 3).

taxon	fraction *de novo* OTUs represented	fraction MAGs represented	FL clusters in SILVA	V4 clusters in SILVA	extant FL based on *de novo*	extant V4 based on *de novo*	extant FL based on MAGs
Bacteria	0.496	0.382	201,586	84,397	406,423	170,155	527,712
Cyanobacteria	0.514	0.375	2,906	1,673	5,654	3,255	7,749

Supplementary Table 7: **Tree distances.** Pairwise topological distances between timetrees considered in the main article and previously published trees (Earth Microbiome Project deblurred 150 bp sequences [10] and SILVA 16S rRNA-based guide tree [22]), using the normalized Robinson-Foulds metric [24]. Also listed are the numbers of tips included in each comparison (i.e., common to both trees compared). See the Methods for details. For a visualization of tree distances see Supplementary Fig. 24.

tree 1	tree 2	distance	tips
Bacteria (16S de novo, FastTree, PATHd8)	EMP deblur	0.82	12,416
Bacteria (16S de novo, FastTree, PATHd8)	Bacteria (16S SILVA, FastTree, PATHd8)	0.82	24,169
Bacteria (16S de novo, FastTree, PATHd8)	SILVA guide	0.84	25,443
EMP deblur	Bacteria (16S SILVA, FastTree, PATHd8)	0.76	28,302
EMP deblur	SILVA guide	0.80	29,930
Bacteria (16S SILVA, FastTree, PATHd8)	SILVA guide	0.74	448,112
Cyan. (16S de novo, FastTree, PATHd8)	EMP deblur	0.73	153
Cyan. (16S de novo, FastTree, PATHd8)	Cyan. (16S SILVA, FastTree, BEAST+PATHd8)	0.71	300
Cyan. (16S de novo, FastTree, PATHd8)	Cyan. (16S SILVA, RAxML, BEAST+PATHd8)	0.71	300
Cyan. (16S de novo, FastTree, PATHd8)	SILVA guide	0.74	326
Cyan. (16S de novo, FastTree, PATHd8)	Cyan. (16S SILVA, BEAST)	0.44	26
EMP deblur	Cyan. (16S SILVA, FastTree, BEAST+PATHd8)	0.71	435
EMP deblur	Cyan. (16S SILVA, RAxML, BEAST+PATHd8)	0.70	435
EMP deblur	SILVA guide	0.74	473
EMP deblur	Cyan. (16S SILVA, BEAST)	0.40	36
Cyan. (16S SILVA, FastTree, BEAST+PATHd8)	Cyan. (16S SILVA, RAxML, BEAST+PATHd8)	0.51	6302
Cyan. (16S SILVA, FastTree, BEAST+PATHd8)	SILVA guide	0.68	6308
Cyan. (16S SILVA, FastTree, BEAST+PATHd8)	Cyan. (16S SILVA, BEAST)	0.68	483
Cyan. (16S SILVA, RAxML, BEAST+PATHd8)	SILVA guide	0.69	6302
Cyan. (16S SILVA, RAxML, BEAST+PATHd8)	Cyan. (16S SILVA, BEAST)	0.69	483
SILVA guide	Cyan. (16S SILVA, BEAST)	0.75	557

S.7 Summary of supplementary files

Supplementary File 1: Accession numbers and summaries of amplicon sequencing data used to recover *de novo* OTUs.
Supplementary File 2: Accession numbers and summaries of sequencing data used from the Earth Microbiome Project[10].
Supplementary File 3: R code used for analyzing diversification dynamics.
Supplementary File 4: Timetrees and undated phylogenetic trees constructed in this study, including required input files.
Supplementary File 5: Taxonomic classifications of *de novo* OTUs.

References

[1] Kendall, D. G. On some modes of population growth leading to ra fisher's logarithmic series distribution. *Biometrika* **35**, 6–15 (1948).

[2] Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **344**, 305–311 (1994).

[3] Maddison, W. P., Midford, P. E., Otto, S. P. & Oakley, T. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**, 701–710 (2007).

[4] Sanderson, M. J. & Donoghue, M. J. Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology & Evolution* **11**, 15–20 (1996).

[5] Morlon, H. Phylogenetic approaches for studying diversification. *Ecology Letters* **17**, 508–525 (2014).

[6] Sanmartín, I. & Meseguer, A. S. Extinction in phylogenetics and biogeography: From timetrees to patterns of biotic assemblage. *Frontiers in Genetics* **7**, 35 (2016).

[7] Norris, J. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, 1998).

[8] Marshall, C. R. Five palaeobiological laws needed to understand the evolution of the living biota. *Nature Ecology & Evolution* **1**, 165 (2017).

[9] Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* (2017).

[10] Thompson, L. R. *et al.* A communal catalogue reveals earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

[11] Schloss, P. D., Girard, R. A., Martin, T., Edwards, J. & Thrash, J. C. Status of the archaeal and bacterial census: an update. *mBio* **7** (2016).

[12] Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences* **113**, 5970–5975 (2016).

[13] Willis, A. Extrapolating abundance curves has no predictive power for estimating microbial biodiversity. *Proceedings of the National Academy of Sciences* **113**, E5096 (2016).

[14] Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biology* **12**, 69 (2014).

[15] Reeder, J. & Knight, R. The 'rare biosphere': a reality check. *Nat Meth* **6**, 636–637 (2009).

[16] Edgar, R. C. Accuracy of microbial community diversity estimated by closed- and open-reference otus. *PeerJ* **5**, e3889 (2017).

[17] Louca, S. & Doebeli, M. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2017).

[18] Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824 (2010).

[19] Liow, L. H., Quental, T. B. & Marshall, C. R. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Systematic Biology* **59**, 646–659 (2010).

[20] Beaulieu, J. M. & O'Meara, B. C. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* **69**, 1036–1043 (2015).

[21] Aldous, D. J. Stochastic models and descriptive statistics for phylogenetic trees, from yule to today. *Statistical Science* **16**, 23–34 (2001).

[22] Glöckner, F. O. *et al.* 25 years of serving the community with ribosomal rna gene reference databases and tools. *Journal of Biotechnology* **261** (2017).

[23] Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **505**, 89–99 (2014).

[24] Day, W. H. E. Optimal algorithms for comparing trees with labeled leaves. *Journal of Classification* **2**, 7–28 (1985).

[25] Jetz, W. *et al.* Global distribution and conservation of evolutionary distinctness in birds. *Current Biology* **24**, 919–930 (2014).

[26] Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S. & Bremer, K. Estimating divergence times in large phylogenetic trees. *Systematic Biology* **56**, 741–752 (2007).

[27] Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J. & Fischer, W. W. Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* **15**, 19–29 (2017).

[28] Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455**, 1101–1104 (2008).

[29] Gibson, T. M. *et al.* Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology* **46**, 135–183 (2017).

[30] Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences* **108**, 13624–13629 (2011).

[31] Brocks, J. J. *et al.* Biomarker evidence for green and purple sulphur bacteria in a stratified palaeoproterozoic sea. *Nature* **437**, 866–870 (2005).

[32] Walter, M., Buick, R. & Dunlop, J. Stromatolites 3,400–3,500 Myr old from the North Pole area, Western Australia. *Nature* **284**, 443–445 (1980).

[33] Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the earth 4.4 gyr ago. *Nature* **409**, 175–178 (2001).