

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The <u>exact sample size</u> (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A full description of the statistics including <u>central tendency</u> (e.g. means) or other basic estimates (e.g. regression coefficient) AND <u>variation</u> (e.g. standard deviation) or associated <u>estimates of uncertainty</u> (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | Clearly defined error bars
<i>State explicitly what error bars represent (e.g. SD, SE, CI)</i> |

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection. All data analyzed were already available on public repositories.

Data analysis

Any custom software used is either included as Supplemental material, or freely available at the CRAN R package repository (<https://cran.r-project.org>). Any 3rd party software used is publicly available and cited in the Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Amplicon sequencing data used to recover de novo OTUs are publicly available under the accession numbers listed in Supplementary File 1. Accession numbers for sequencing data used from the Earth Microbiome Project are listed in Supplementary File 2. R code used in this study is provided as Supplementary File 3. Timetrees

as well as undated phylogenetic trees constructed in this study are provided as Supplementary File 4. Taxonomic classifications of de novo OTUs are provided as Supplementary File 5.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We reconstruct diversification dynamics of bacteria using publicly available 16S rRNA gene sequence data and novel phylogenetic methods.
Research sample	All analyses were based on existing, publicly available DNA sequence data. Accession numbers are provided as Supplementary Material.
Sampling strategy	Amplicon sequencing data, used to generate our de-novo trees, were chosen so as to represent as wide of an environmental range as possible, under the constraint that they must cover at least 200 bp of the 16S V4 region. For the SILVA trees, all OTUs available within the appropriate taxon were used. For the BEAST-computed trees, trees were sub-sampled randomly to enable computational feasibility. Any sub-sampling has been accounted for in our calculations.
Data collection	No novel data was collected.
Timing and spatial scale	No novel data was collected.
Data exclusions	No relevant data was explicitly excluded.
Reproducibility	All raw sequencing data used are available at public repositories under accession numbers provided as Supplemental material. SILVA 16S sequences are publicly available at the SILVA project website (https://www.arb-silva.de). All trees created in this study, and computer code needed to analyze them, are provided as supplemental material. All other software used in our analyses are publicly available and cited in our Methods section.
Randomization	This study does not include experimental groups.
Blinding	This study does not include experimental groups.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

Materials & experimental systems

- n/a Involved in the study
- Unique biological materials
 - Antibodies
 - Eukaryotic cell lines
 - Palaeontology
 - Animals and other organisms
 - Human research participants

Methods

- n/a Involved in the study
- ChIP-seq
 - Flow cytometry
 - MRI-based neuroimaging