

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | 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> |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | 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> |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

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

Data collection No software was used for data collection.

Data analysis Custom scripts have been deposited at Figshare (<https://figshare.com/s/59b28ecc0056dc8d0d03>). Published software used for data analysis are described in the Methods section, and include: DIAMOND v0.9.29/v0.9.36, HMMER v3.3, CD-HIT v4.7, MAFFT v7.471, trimAl v1.4, BMGE v1.12, IQ-TREE v1.6.12/v2.0.3, Phylobayes v1.8c, Treemmer v0.3, GFMix v1.1 (<https://www.mathstat.dal.ca/~tsusko/software.html>), MEOW v0.6.4 (<https://github.com/RogerLab/meow>), and R v4.2.3

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequence alignments, phylogenetic trees, and custom scripts are available at Figshare (<https://figshare.com/s/59b28ecc0056dc8d0d03>) and the accession codes for publicly available data used in these analyses are given in Supplementary Table 7. Databases in which the data with the accession codes can be found are at the National Centre for Biotechnology Information (NCBI: <http://www.ncbi.nlm.nih.gov>), the MMETSP database located at the iMicrobe site <https://www.imicrobe.us/#/projects/104> and the Phylofisher v1 dataset (<https://doi.org/10.6084/m9.figshare.15141900.v1>)

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

NA

Reporting on race, ethnicity, or other socially relevant groupings

NA

Population characteristics

NA

Recruitment

NA

Ethics oversight

NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size

Sampling of genes was based on Munoz-Gomez et al. (2022) Nature Ecol. Evol. 6:253-262. Sampling of taxa was done to maximize phylogenetic diversity within eukaryotes and coverage of the genes chosen but downsampled to allow computational feasibility of the analyses. Detailed information about how the sampling was conducted is given in the Methods section.

Data exclusions

No data was excluded

Replication

In maximum likelihood analyses, bootstrapping and approximate likelihood ratio tests were used to evaluate the robustness of the results. In Bayesian analyses, posterior probabilities were used.=

Randomization

Randomization is not necessary in studying phylogenetic relationships as these analyses use objective algorithmic approaches to infer evolutionary relationships from fixed data (i.e., protein sequence data).

Blinding

Blinding is not necessary in phylogenetic analyses as they rely on objective computational methods that use predefined algorithms to infer evolutionary relationships from sequence data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involvement in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

NA

Novel plant genotypes

NA

Authentication

NA