

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 ncbi-genome-downloader (v0.3.3)

Data analysis

Multiple software packages were used for genomic and evolutionary analyses, including: IQ-TREE(v1.6.12 & v2.2.0.3), antiSMASH (v7.0.0), MCMCtree (v4.10.0), I1ou (v1.43), HMMER (v3.3.2), SandPiper (v0.2.0), OrthoFinder (v2.5.4), MUSCLE (v5.1), trimal (v1.4.1), funannotate (v1.8.15), UFGC profile (v1.0.5), pyrodigal (v3.0.0), geNomad (v1.11.0), pyhmmmer (v0.10.4), CD-HIT (v4.8.1), DIAMOND alignment (v2.0.15), Pyseer (v1.3.11), SplitsTree4CE (v4.19.2), BLASTn (v2.5.0), FastTree2 (v2.1.11), GToTree (v1.7.07), synthasr (v1.1.22).

We also used programs we developed ourselves such as the zol suite (v1.3.9, v1.3.11, & v1.6.13) and codoff (v1.2.3) and new programs we developed specifically for this manuscript: psaps (v1.0.0) and paiof (v1.0.2b). All custom scripts not provided as independent software are provided in a separate GitHub directory. All GitHub repositories have been released. Releases for them are also tracked in immutable forms on Zenodo to ensure reproducibility.

GitHub and Zenodo links to bioinformatic tools we developed or significantly improved specifically for this manuscript:

codoff = <https://github.com/Kalan-Lab/codoff> ; <https://zenodo.org/records/17585739>

psaps = <https://github.com/raufs/psaps> ; <https://zenodo.org/records/15677124>

paiof = <https://github.com/raufs/paiof> ; <https://zenodo.org/records/17944403>

Additional scripts and code associated specifically with this manuscript are provided on GitHub at: https://github.com/Kalan-Lab/Ancestral_BGC_Extensions_Study. Releases of this repo are also stored on Zenodo (latest: <https://zenodo.org/records/19632677>)

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](#)

NCBI accession codes for genomic assemblies used for different analyses presented in this study are provided in Supplementary Table S1, S3, S4, S6, S7, and S11. Larger datasets, including select files from antiSMASH BGC predictions, and fungal gene calling data can be found on the Zenodo repository: <https://zenodo.org/uploads/15168076>.

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

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

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)

Ecological, evolutionary & environmental sciences study design

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

Study description

Comparative genomics, evolutionary, and phylogenetic analyses were performed using microbial genomes downloaded from public genomic databases. Selection of representative genomes were performed for individual taxonomic groups based on assembly contiguity.

Research sample	Vetted bacterial genomes for contamination and completeness from the Genome Taxonomy Database (GTDB release 214) were downloaded from NCBI. Representative genomes for species, genera, and orders were selected based on assembly quality. Fungal genomes were selected using a similar approach from the Universal Fungal Core Genes (UFCG) database. To sample early-diverging fungi, additional genomes from the study by Amses et al. 2022 were also included.
Sampling strategy	Taxon representative genomes were selected to avoid database bias, since some species are highly sequenced, such as lineages which are frequently sequenced from nosocomial settings.
Data collection	Data were downloaded from NCBI RefSeq and GenBank databases, with details, including accessions noted in supplementary tables.
Timing and spatial scale	Genomes were downloaded in mid to late 2023.
Data exclusions	For some analyses, genomes with low N50, an assembly contiguity statistic, were disregarded since in tact biosynthetic gene clusters are more difficult to detect in them and could cause technical artifacts in quantitative investigations/associations.
Reproducibility	For major claims made in the study, such as our proposition that the last common ancestor of the class Actinomycetes was multicellular, multiple independent analyses were employed to provide rigor to claims. For large scale assessment of trends - multiple distinct lineage representatives were used and distributions of metrics assessed. Our analyses were performed using conda environments with information on major software and dependencies tracked. Releases for all three software GitHub repos and the main GitHub repo associated with this study are also tracked on Zenodo.
Randomization	Genomes were grouped by taxonomic membership according to GTDB (for bacteria) or NCBI (for fungi).
Blinding	Since data were gathered from public sources it was not possible to apply blinding.
Did the study involve field work?	☐ Yes ☒ No

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	Involved 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	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	No plant experiments were performed
Novel plant genotypes	No plant experiments were performed
Authentication	No plant experiments were performed