nature portfolio

Corresponding author(s): Baolan Hu

Last updated by author(s): Dec 4, 2023

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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	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 *Give P values as exact values whenever suitable.*
☒	☐	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	ZEN 2012 blue edition (Zeiss, Germany), SerialEM (v.3.9), Diamond (v.0.9.18), Basic Local Alignment Search Tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi), Proteome Discoverer (v.2.4), Integrated Microbial Next Generation Sequencing (IMNGS) platform (http://www.immgs.org/).
Data analysis	OriginPro (v.8.5), PreSens Measurement Studio (v.2.0), ZEN 2012 blue edition (Zeiss, Germany), IMOD (v.4.11), Tomo3D (v.2.1), Trimmomatic (v.0.36), MEGAHIT (v.1.0.6, http://github.com/voutcn/megahit), MetaGeneMark (v.3.38, http://exon.gatech.edu/GeneMark/metagenome/Prediction), CD-HIT (v.4.7), BBMAP software (http://jgi.doe.gov/data-and- tools/bbtools/), MetaErg (https://github.com/xiaoli-dong/metaerg), MetaBAT (v.2.12.1), CheckM (V.1.0.11), GTDBTk (http://github.com/Ecogenomic/GtdbTk), MEGA 7 (v.1.0.0), iTOL (v 6), Adobe Illustrator CC 2018, SortMeRNA (http://bioinfo.lifl.fr/RNA/sortmerna), Bowtie2 (v.2.33, https://github.com/BenLangmead/bowtie2), corset (v.1.06), ArcGIS 10.5.

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

The following databases/datasets were used in this study: GTDB (v2.2.1, https://github.com/Ecogenomics/GTDBNCBI), NCBI (https://www.ncbi.nlm.nih.gov/), SRA (https://www.ncbi.nlm.nih.gov/sra) and KEGG (http://www.kegg.jp/kegg/).
Raw data of the 16S rRNA gene sequencing were submitted to the Sequence Read Archive (SRA) with accession numbers of SRR21143259-SRR21143272, SRR23318916-SRR23318920. The metagenomic and metatranscriptomic sequencing data and MAGs generated in this study were deposited in NCBI database under BioProject number PRJNA869304. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD047070. Representative images of FISH and microscopy have been deposited in Figshare. Source data are provided with this paper.

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	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine the sample size. The sample size used in this study were based on past experiences in the filed of research (Zhou, Z. et al., 2021, Nature; Garrido-Amador, P. et al., 2023, Nature Microbioloy; Zehnle, H. et al., 2023, Nature Microbiology). Within current study, a moderate number (n=3) of biologically independent replicates were used for key experiments, such as isotope labeling batch tests, transcriptomics, proteomics and FISH of the highly purified culture etc., which were more crucial for the conclusions of this study. Given that excessive sampling from reactors would negatively affect the cultivation of 'M. sinica' due to its extremely low growth rate, two or one biological replicates were set for those regular monitoring tests or less important experiments, such as the 16S rRNA gene amplicon sequencing and quantitative PCR of enrichment cultures etc. In conclusion, the sample size in this study is a balance, based on previous experiences, between enough replicates to get statistical power and sampling management.
Data exclusions	No data were excluded from the analysis.
Replication	1) qPCR: Two biological replicates from the enrichment stage (day 0, 30, 100, 200, 260, 290, 320, 340, 360 and 380). 2) 16S sequencing: Two biological replicates from the enrichment stage (day 0, 100, 200, 290, 340, 360 and 380). One biological replicate from the purification stage (day 410, 780, 870, 960 and 1330) (due to low biomass). 3) FISH: Three biological replicates from the enrichment stage (day 380). And three biological replicates from the the purification stage (day 1330). 4) Isotope labeling experiments: Three biological replicates from the the purification stage (day 1330). 5) Transmission electron microscopy and scanning electron microscopy: Three biological replicates from the enrichment stage (day 380). 6) Phase-contrast microscopy and Cryo-ET: Three biological replicates from the purification stage (day 1330). 7) Metagenomics: One biological replicate from the enrichment stage (day 290, 340, 360 and 380).

8) RNA-seq: Three biological replicates from the purification stage (day 1330). Two biological replicates from the enrichment stage (day 380). And one biological replicate from the enrichment stage (day 290, 340 and 360).
9) Proteomics: Three biological replicates from the enrichment stage (day 380). And three biological replicates from the purification stage (day 1330).
10) Substrate affinity tests: Three biological replicates from the purification stage (day 1330).
11) Oligotrophic and copiotrophic incubations: Two biological replicates from the enrichment stage (day 360).
12) Chemical analyses: Two biological replicates from the enrichment stage (day 0, 30, 100, 200, 260, 290, 320, 340, 360 and 380).
All attempts at replication were successful.

Randomization

This is an exploratory study targeting a novel microbial process and there are no experimental groups set for a lone-term cultivation. All experiments were performed using an enrichment culture or a highly purified culture. Intrinsically all samples taken from the bioreactor are randomized because every cell has the same statistical probability to be sampled. In addition, as for the nitrate and nitrite affinity tests, all other experimental conditions (temperature, time, concentration, etc.) were controlled to assure the validity of the results.

Blinding

No blinding was done in this study. All tests were based on anaerobic cultures in the lab, which required us keeping careful track of all conditions and monitoring the growth, therefore blinding such studies seemed difficult.

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