nature portfolio

Corresponding author(s): Yu-Rong Liu (yrliu@mail.hzau.edu.cn)

Last updated by author(s): May 13, 2025

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	The data for this manuscript were collected from lab analyses of both paddy and upland soils (0-15 cm depth) from the two typical agricultural regions in Wanshan, Southwest China, and Xianning, Center China. Measurements of CH3Hg+ concentration, soil properties and soil microbial communities in microcosm experiments and pure culture assays were conducted as described in the Method section of our manuscript. All code associated with our analyses in this study is available at https://figshare.com/s/fd35576a3ffb97ea098b.
Data analysis	Raw data were retrieved and filtered with the Circular Consensus Sequencing (CCS) software of the SMRT Link (minfullpass = 3, polish min Predicted Accuracy = 0.8, minLength = 500) to obtain the Raw CCS Reads. The barcode reads of every sample were recognized with Lima (https://github.com/PacificBiosciences/barcoding) to acquire Raw CCS, whereafter the CCS (accuracy above 99%) of each sample was tested and chimeric reads were removed with UCHIME (http://drive5.com/uchime, v8.1) to acquire quality controlled sequences. Subsequently, the representative sequences were compared using the Mothur (https://mothur.org/wiki/Classify.seqs, v1.40.0) software with Silva 132 (https://www.arb-silva.de/documentation/release-132/) database (classified at a bootstrap threshold of 0.9) to gain classified information. A total of 10.7-18.6 Gb clean reads were obtained for each sample after quality control following the KneadData pipeline (v0.6.1), accounting for 94.3%-97.9% of raw bases. De novo assembly of clean reads into contigs was performed using megahit (https://github.com/voutcn/megahit, -k-min 47 -k-max 97 -k-step 10, v1.2.9), CD-HIT (http://www.bioinformatics.org/cd-hit/, v4.6.1) and Salmon (https://github.com/COMBINE-lab/salmon, v1.10.0) were employed for subsequent open reading frames (ORFs) prediction, non-redundant gene catalogue clustering, and non-redundant gene quantification. Biological function annotation of the non-redundant genes was conducted on GhostKOALA annotation servers (Automatic KO assignment by GHOSTX sequence similarity search. Binning of the assembled metagenome was computed and refined using MaxBin2 (http://sourceforge.net/projects/maxbin/, v2.2.7) and metaBAT2 (https://bitbucket.org/berkeleylab/metabat, v2.12.1) in MetaWRAP. Completion and contamination estimation of all the bins were computed by CheckM (https://github.com/Ecogenomics/CheckM/wiki, v1.1.6). Taxonomic classification of MAGs was performed using GTDB-Tk v0.3.3 classify_wf command against the GTDB (http://gtdb.ecogenomic.org/, v2.3.2). The predicted genes of MAGs were annotated using the HMM profile database for KEGG orthology

with predefined score thresholds through KofamScan (v1.3.0).
One-way analysis of variance (ANOVA) followed by two-sided Tukey post hoc test was performed by SPSS Statistics 21.0 (IBM, Chicago, IL) to detect the significant differences between treatments in the pure culture assays. LEfSe analysis ($p < 0.05$, LDA score > 2) was used to identify bacterial biomarkers for the 13CH3Hg+ and 12CH3Hg+ treatments on the open website (http://huttenhower.sph.harvard.edu/galaxy). A two-sided Student's t-test was conducted to identify genes or microbial species that exhibited significant enrichment in the heavy fraction of the 13C-labeled treatment compared to the corresponding fraction in the 12C control (* representing $p < 0.05$), and to assess differences in CH3Hg+ degradation efficiency between paddy and upland soils at both high and low Hg-contaminated sites. The phylogenetic trees were generated and visualized using a CVTree4 algorithm (https://cvtree.online/v4/prok/) and the Evolview v3 web application (https://www.evolgenius.info/evolview/#/treeview), respectively.

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 full-length 16S rRNA gene amplicon sequences and metagenomic sequences generated in this study have been deposited to the NCBI SRA database under the BioProject IDs of PRJNA1183541 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1183541) and PRJNA1183649 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1183649), respectively. The bioaccumulation factor and rice ingestion rate used in this study were collected from https://doi.org/10.1038/s43016-024-00954-7 and https://doi.org/10.1038/s41467-019-09080-6, respectively. The databases used in this study include SILVA 132 (https://www.arb-silva.de/documentation/release-132/), Genome Taxonomy Database (http://gtdb.ecogenomic.org/) and KEGG database (https://ww.genome.ad.jp/kegg/). 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

Ecological, evolutionary & environmental sciences study design

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

Study description	This study aims to explore specific microbial taxa and metabolic pathways involved in CH3Hg+ degradation and to estimate the microbial contribution to CH3Hg+ mitigation in rice. We employed 13CH3Hg+-DNA stable-isotope probing combined with 16S rRNA long-read sequencing to identify microbial taxa responsible for 13CH3Hg+ degradation, and subsequently validated the degradation ability of putative taxa through pure culture assays. We further reconstructed metagenome-assembled genomes to explore potential metabolic pathways associated with CH3Hg+ degradation. Finally, we estimated potential changes in rice CH3Hg+ accumulation and associated health risks resulting from microbial degradation of soil CH3Hg+ in the typical paddy soils from main rice-producing areas in China.
Research sample	Both the paddy and upland soils (0-15 cm depth) were collected from the two typical agricultural regions: Wanshan in Southwest China and Xianning in Central China. Wanshan is known as the "Hg hotspot" of China, with more than 600 years of Hg mining and smelting activities, leading to elevated Hg concentrations in these soils. While there are no documented Hg pollution sources in Xianning, Hg in the soil is most likely derived from parent materials and atmospheric deposition. Thus, the two sites represent high

and low Hg contamination scenarios, respectively. Paddy and upland rotations are used as common cropping systems in both regions, so the presence of CH3Hg+ in the soil in both areas should not be disregarded. These soils were used in a 13CH3Hg+ degradation experiment to assess CH3Hg+ degradation potential. Soil CH3Hg+ and chemical properties were analyzed for all four soil types (high- and low-Hg paddy and upland soils).

Considering the relatively high cost of DNA-SIP experiments and the need for sufficient 13C incorporation to ensure accurate identification of key microbial taxa, we selected the paddy soil from high Hg-contaminated site and the upland soil from low Hg-contaminated site, which exhibited higher microbial degradation potential, for subsequent 13CH3Hg+-DNA-SIP analyses. These representative soils were chosen to reflect two contrasting Hg contaminated scenarios (high vs. low Hg) under different land-use types (paddy vs. upland), and to provide insights into the microbial taxa involved in CH_3Hg^+ degradation across different environments.

To validate the ability of candidate microbial taxa to degrade CH3Hg+, we conducted a series of bioassays using Dechloromonas denitrificans (DSM 15892) and Methylovorus menthalis (DSM 24715), which belong to the proposed genera associated with CH3Hg+ degradation. Dechloromonas denitrificans (DSM 15892) and Methylovorus menthalis (DSM 24715) were purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany).

Sampling strategy

Both the paddy and upland soils (0-15 cm depth) were collected from the two typical agricultural regions: Wanshan in Southwest China and Xianning in Central China. At each site, we established five 1 m × 1 m sub-plots positioned at each corner and center of a 50 m × 50 m area. Five soil cores (top 15 cm) were collected and then mixed into a composite soil sample to consider the heterogeneity. This approach enabled us to capture the variability in soil properties across the study area.

We did not perform a formal statistical method to predetermine sample size. This is because the primary objective of the study was to identify key microbial taxa and metabolic pathways involved in CH3Hg+ degradation using 13CH3Hg+-DNA-SIP analyses and metagenomics, both of which are time-consuming and highly costly. Therefore, representative soils with high CH3Hg+ degradation potentials were selected based on preliminary degradation assays. The sample sizes used are consistent with previous SIP-based studies and considered sufficient to identify the dominant active degraders in soil.

Data collection

The original data of this study were collected mainly by Xin-Quan Zhou, Kang-Hua Chen, Man Yang, Qin Liu, Yun-Yun Hao and Hui-Wen Liu. The concentration of CH3Hg+ was measured by an Automated Methyl Mercury Analytical System (MERX) (Brooks Rand Labs, Seattle, WA, USA). Soil pH was measured by a Delta pH-meter (Mettler-Toledo Instruments Co., Columbus, OH, USA) in a soil slurry with a water-to-soil ratio of 2.5:1. Dissolved C in soils was extracted with 0.5M K2SO4, and the extracted solution was used for analyses of DOC by a total organic C analyzer (vario TOC; Elementar). Total genomic DNA was extracted using a MoBio power soil DNA isolation kit (MO BIO Laboratories, Inc., Carlsbad, CA) according to the manufacturer's instructions. DNA concentrations were determined using an ND-2000 UV-vis spectrophotometer (NanoDrop Technologies, Wilmington, DE). The buoyant density value of each DNA fraction across CsCl gradients was measured using a digital refractometer (Reichert AR2000). Amplification of full-length (V1–V9) 16S rRNA gene sequences was performed on a PacBio R RS II platform (Majorbio Bio-Pharm Technology, Co., Ltd., Shanghai, China). The metagenomic shotgun sequencing (2 ×150 bp) was carried out on an Illumina NovaSeq 6000 platform (Shanghai Majorbio Bio-pharm Technology). The bioaccumulation factor and rice ingestion rate used in this study were collected from https://doi.org/10.1038/s43016-024-00954-7 and https://doi.org/10.1038/s41467-019-09080-6, respectively.

Timing and spatial scale

Both paddy and upland soils (0-15 cm depth) were collected in 2018 (from September to November) from two typical agricultural regions: Wanshan in Southwest China (27.52°N, 109.21°E) and Xianning in Central China (29.90°N, 114.23°E). The 13CH3Hg+ degradation experiment was sampled on days 0, 3, 7, 14 and 28. The DNA-SIP experiment was sampled on days 0, 7, 14 and 28.

Data exclusions

No data were excluded in the analyses.

Reproducibility

Within the manuscript, we clearly state all the steps taken to ensure the reproducibility of the study. We include descriptions of standard sampling and analytical protocols and the identification of all code packages used. All microcosm experiments were conducted with three biological replicates. The pure culture assays were performed twice, and all attempts to repeat the experiment were successful.

Randomization

The samples were not randomly assigned, as the experimental design was intended to compare the treatments across representative soils with different Hg contamination levels and land-use types.

Blinding

Blinding was not applied, as the data collection and analysis were based on objective measurements (e.g., chemical concentrations, sequencing results) obtained using standardized protocols and automated instruments.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions

Both paddy and upland soils were collected from the two typical agricultural regions in Wanshan, Southwest China, and Xianning, Center China. The mean annual precipitation and temperature in these areas ranged from 1378.7-1577.4 mm and 13.7-16.8°C, respectively. More details can be found in the Methods section of the manuscript.

Location

Both paddy and upland soils were collected from two typical agricultural regions: Wanshan in Southwest China (27.52°N, 109.21°E) and Xianning in Central China (29.90°N, 114.23°E). Detailed information on the geographical coordinates of sampling locations included in this study was presented in the Supplementary Table. 1 and Methods.

Access & import/export

All soil samples collected by our co-authors in full compliance with local law, and no permission is needed for sample collection.

Disturbance

This study did not cause any environmental disturbance.

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	N/A
Novel plant genotypes	N/A
Authentication	N/A