Supporting Information

Microbial communities associated with methylmercury degradation in paddy soils

Xin-Quan Zhou,[†‡] Yun-Yun Hao,[‡] Baohua Gu,[§] Jiao Feng,[‡] Yu-Rong Liu, [*†‡ǁ] Qiaoyun Huang, [†‡ǁ]

[†]State Key Laboratory of Agricultural Microbiology and [‡]College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China

[§]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

[ǁ]Hubei Key Laboratory of Soil Environment and Pollution Remediation, Wuhan, 430070, China.

* **Corresponding author:**

Dr. Yu-Rong Liu. College of Resources and Environment, Huazhong Agricultural University, Wuhan, 430070, China. E-mail: yrliu@mail.hzau.edu.cn; Phone: (+86) 27-87676165

(25 pages, 1 text, 8 tables, 13 figures)

Contents

One Text:

Eight Tables:

Thirteen Figures:

Text S1: Methods of total RNA extraction and metatranscriptomic analysis

Total mRNA was extracted from 2 g of soil by using RNeasy® PowerSoil® Tatal RNA Kit (MoBio Laboratories, Carlsdad, CA, USA), and RNase-free DNase was employed to remove the contaminating DNA before RNA sequence. The concentration and quality of RNA were checked using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The metatranscriptomic sequencing was performed on the Illumina HiSeq platform.

Poor-quality bases of raw reads, reads with "N" bases, and those with lengths of < 50 bp were discarded using were Sickle. [1] Then, the rRNA reads were removed in a BLAST search of the SILVA short subunit (SSU; 16S/18S) and SILVA long subunit (LSU; 23S/28S) database using the SortMeRNA software. [2] Putative mRNA reads were assembled *de novo* using the Trinity software,[3] and characterized with BLASTX comparisons against the integrated National Center for Biotechnology Information (NCBI) nr (nonredundant) protein database (E-values < 10^{-5}).[4] The LCA-based algorithm implemented in MEGAN was used to determine the taxonomic level of each gene. Then open reading frames (ORFs) were predicted based on the assembled contigs using the TransGeneScan software.[5] All the ORFs were clustered using the CD-HIT software (95% identity, 90% coverage) to establish the non-redundancy gene set.[6] BLASTP [7] (the BLASTP program at the E value of 10^{-5}) was used to query the predicted proteins sequence reads against the integrated nr protein database. Functional annotation was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations. The transcript expression was normalized using fragment per kilobase of transcript per million mapped (FPKM).[8]

We did not find the *merA* and *merB* genes by searching in the KEGG annotations directly. Thus, hidden Markov models (HMMs) were used to search metatranscriptomic data for MerA and MerB homologs using the HMMER package. [9] A profile HMM of the entire MerA sequence was generated from 252 representative MerA sequences from NCBI RefSeq database. The MerA sequences were obtained by merging and translating the raw reads sequence, then these protein sequences were searched using HMM. The obtained MerA homologs were validated by BLASTN searches with an E-value cutoff of 1×10^{-5} against the

nr nucleotide database maintained by the NCBI. Similarly, a profile HMM for MerB was generated from 334 known MerB sequences.

Table S1. Site information and initial soil geochemical properties from Hunan, Guizhou and Hubei provinces.

Sites	Longitude (E)	Latitude (N)	pH	NH_4^+ (mg kg^{-1})	NO_3^- (mg kg^{-1})	MeHg (μg·kg^{-1})	THg (μg·kg^{-1})
Hunan	111.48893	28.90259	5.5 ± 0.07	17.5 ± 0.35	6.6 ± 0.82	1.01 ± 0.02	73.58 ± 3.62
Guizhou	109.21369	27.51796	6.9 ± 0.34	1.47 ± 0.74	70.92 ± 6.42	2.03 ± 0.01	2911.52 ± 2051.07
Hubei	109.48817	30.27217	7.44 ± 0.06	3.09 ± 0.09	215.84 ± 26.84	0.35 ± 0.04	59.76 ± 6.12

Note: The mean values and standard deviations are presented.

Table S2. Primer sets and amplification programs used in the quantitative PCR.

Gene	Primer	Sequence (5'-3')	Protocol	Ref
hgcA	Delta-F	GCCAACTACAAGMTGASCTWC	94 °C 3 min; 95 °C 15 s, 65 °C 10 s, 65 °C 10 s, 40 cycles; 72 °C 4min	10
	Delta-R	CCSGCNGCRCACCAGACRTT		
dsrA	DSR1F	ACSCACTGGAAGCACG	95 °C 40 s; 94°C 1 min, 56°C 40 s, 72°C 2 min, 40 cycles; 72°C 4 min	11
	DSR4R	GTGTAGCAGTTACCGCA		
dsrB	DSRp2060F	CAACATCGTYCAYACCCAGGG	95°C 30 s; 95°C 5 s, 55 °C 34 s, 72°C 34 s, 40 cycles; 72°C 4 min	12
	DSR4R	GTGTAGCAGTTACCGCA		
mcrA	Mlas	GGTGGTGTMGGDTTCACMCARTA	95°C 3 min; 95°C 30 s, 55°C 34 s, 72°C 34 s, 40 cycles; 72°C 4 min	13
	McrA-R	CGTTCATBGCGTAGTTVGGRTAGT		
pmoA	A189f	GGNGACTGGGACTTCTGG	94°C 2 min; 94°C 45 s, 53°C 1 min, 72°C 2 min, 40 cycles; 72°C 4 min	14, 15
	Mb661r	CCGGMGCAACGTCYTTACC		

Table S3. The methanotrophic genera in the Cu addition soil based on 16S rRNA sequencing analysis.

	Genera	Relative abundance (%)
Type I	*Methylomonas*	~ 0.02
	Methylosarcina	~ 0.003
	Methylobacter	~ 0.02
	Methylocaldum	~ 0.01
	Methylococcus	~ 0.003
Type II	*Methylocystis*	0.1 ~ 0.3

Table S4. Correlation coefficients (Pearson's ρ) between MeHg degradation (%) and the ecological modules (Mod), identified based on microbial co-occurrence network using default parameters from the interactive platform Gephi.

	Mod 0	Mod 1	Mod 2	Mod 3
MeHg Degradation (%)	-.232*	-.187	-.135	0.265**

Note: * represents $p < 0.05$; ** represents $p < 0.01$

Table S5. Random forest analysis identifying remaining genera of Mod 3 associated with MeHg degradation, in addition to those showed in Figure 3b.

Genera	%IncMSE
Isosphaera	15.525
Sporichthya	15.491
Clostridium_sensu_stricto_10	15.453
uncultured_ge	14.563
Variibacter	8.335
Rhodanobacter	7.786
Singulisphaera	7.754
uncultured	6.247
Oryzihumus	4.774

Table S6. The microbial taxa associated with MeHg degradation in the Hunan paddy soil.

Families	Average relative abundances (%)
Mycobacteriaceae	0.68
Frankiaceae	0.35
Thermomonosporaceae	0.04
Catenulisporaceae	0.24

Table S7. The selected, differentially expressed genes are provided as a separate excel file.

Table S8. Number of MerA and MerB sequences identified using hidden Markov models (HMMs).

Samples	Reads	
	MerA	MerB
T_0	565	3
T_{48}	471	2

Figure S1. Methylmercury **(**MeHg) degradation (%) at 48 h in these paddy soils from Hunan, Guizhou and Hubei provinces.

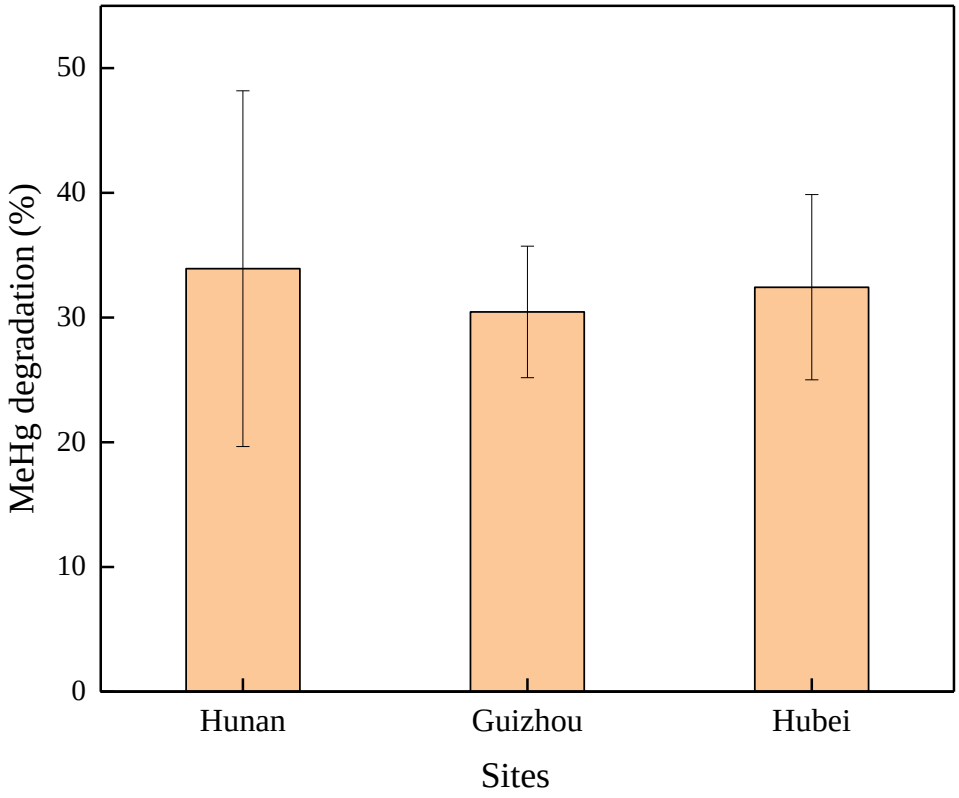


Figure S2. The mercury (Hg) transformation related genes abundance (a) and relative abundances of main phyla in the sterilized soil (b). SRB (*dsrA* and *dsrB*), methanogens (*mcrA*) and methanotrophs (*pmoA*) have been implicated in MeHg degradation.

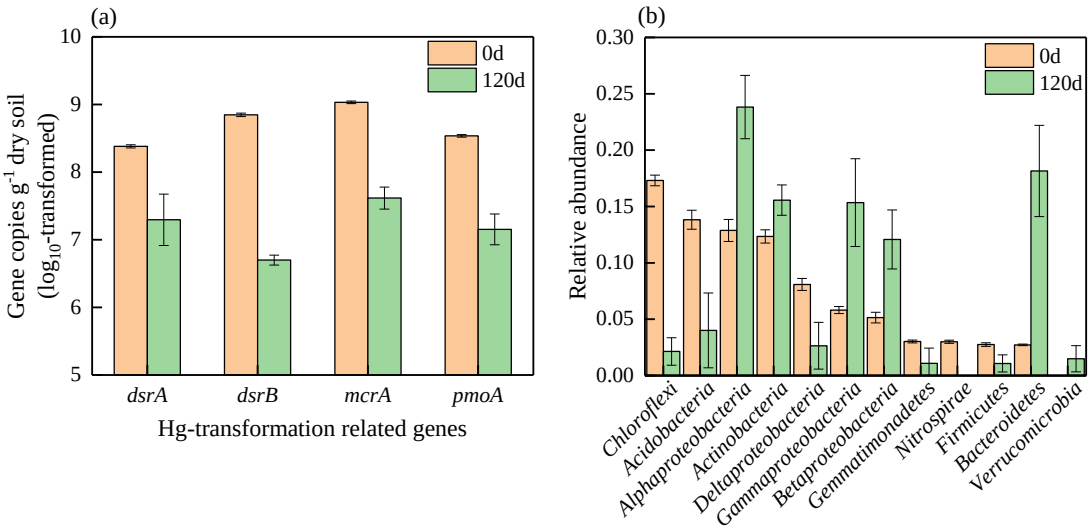


Figure S3. The Hg-transformation related genes abundances in the un-sterilized (CT) and sterilized soils.

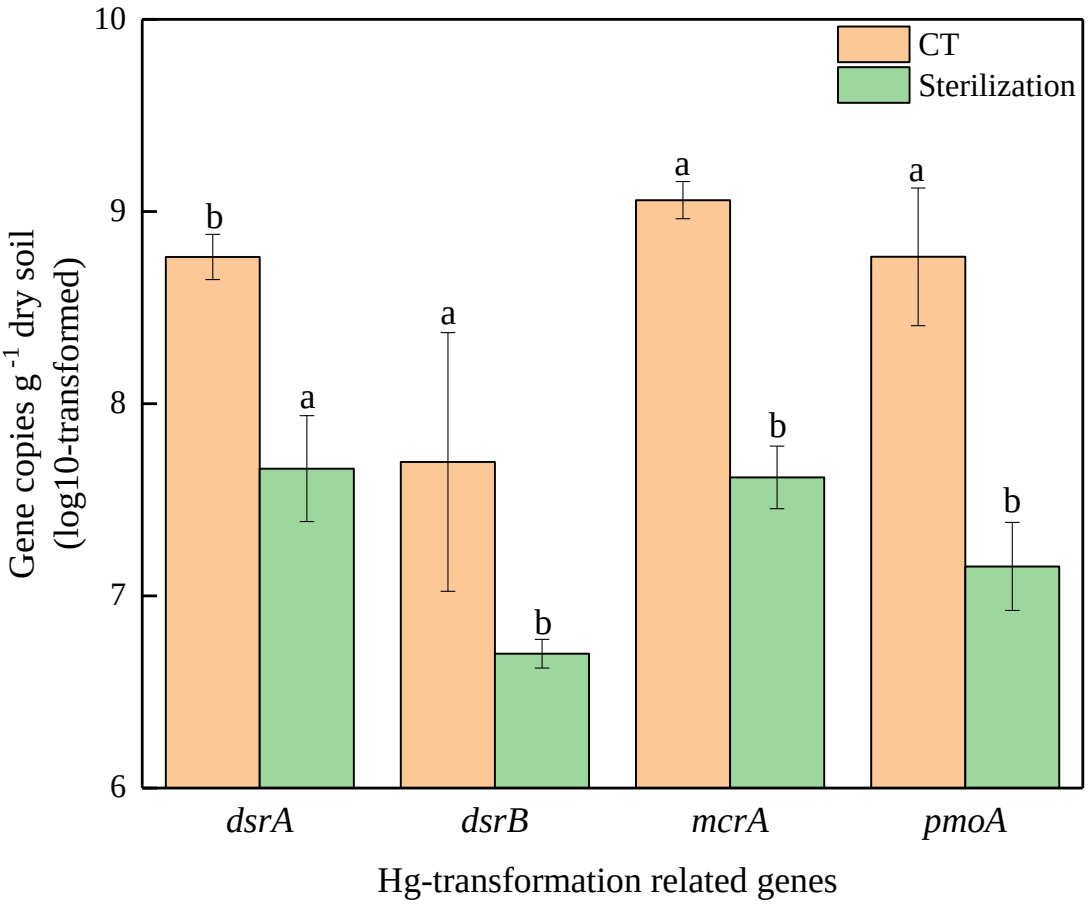


Figure S4. Linear discriminant analysis effect size (LEfSe) for microbial taxa biomarkers at the phylum level under water unsaturated conditions at 120 d. CN represents no MeHg addition, and CT represents spiked with MeHg.

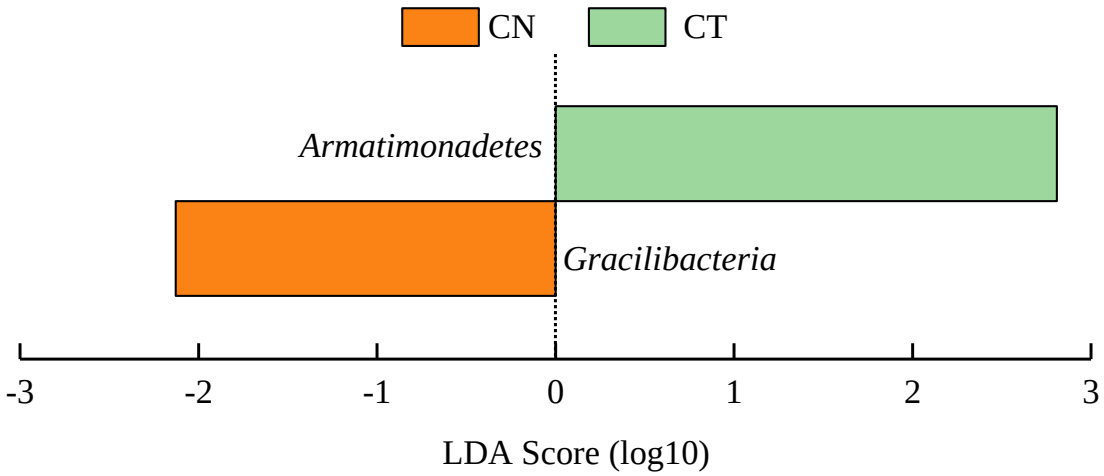


Figure S5. The effect of Cu addition on soil microbial community composition. (a) relative abundance of dominant bacteria at the phylum level, and (b) relative abundance heatmap of the main genera.

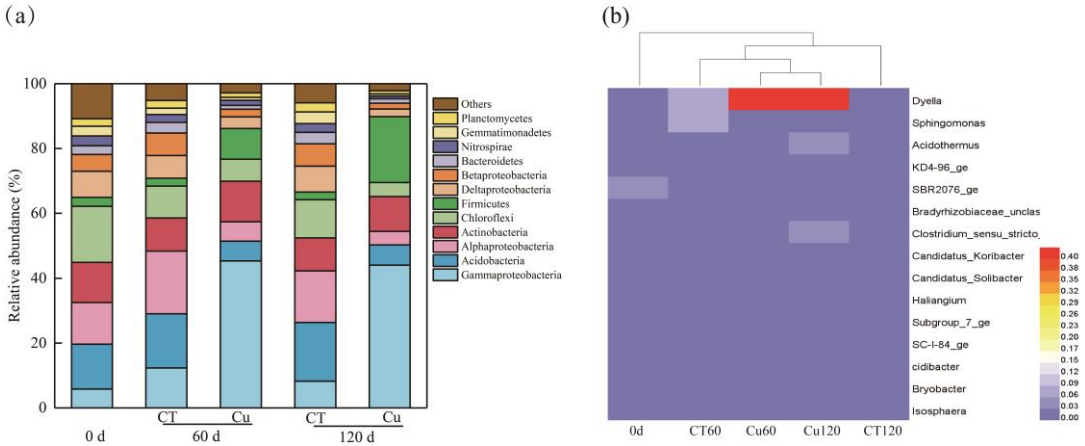


Figure S6. The effects of (a) Cu, (b) molybdate (MoO_4) and (c) bromoethanesulfonic acid (BES) on the abundances of functional genes related to Hg transformation at 0, 60 and 120 d.

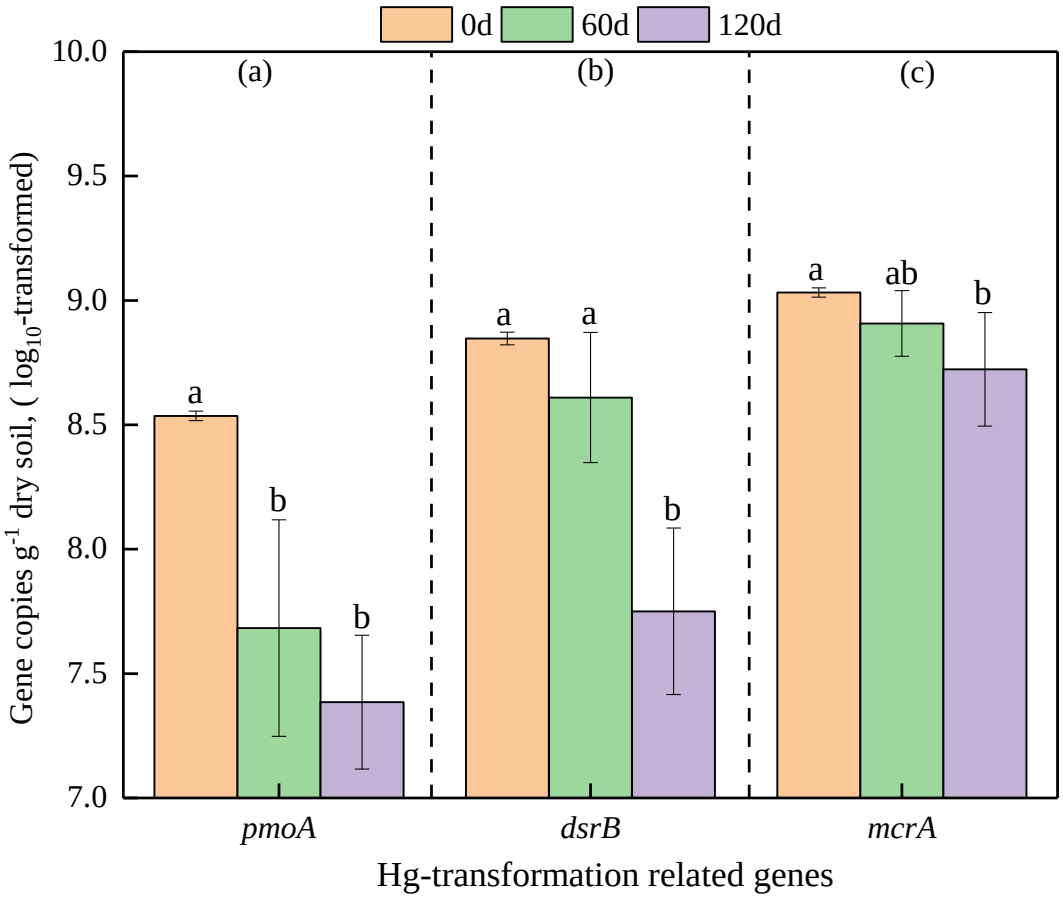


Figure S7. The effects of MoO_4 and BES addition on microbial community composition at the (a) phylum and (b) genus level. CT_W represents control of the water-saturated condition.

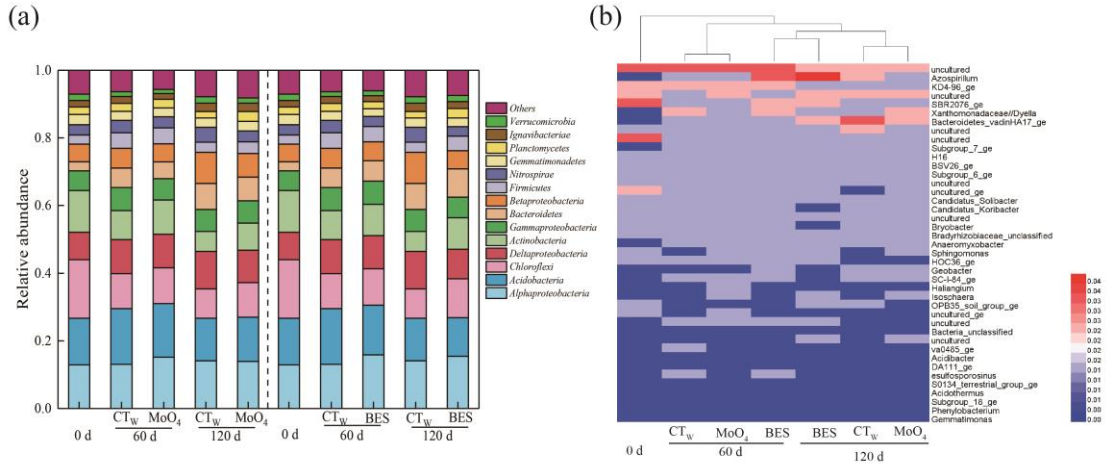


Figure S8. Changes in copy numbers of functional genes related to Hg transformation at 0, 60 and 120 d. NH_4 represents added NH_4Cl treatment; NO_3 represents added $NaNO_3$ treatment.

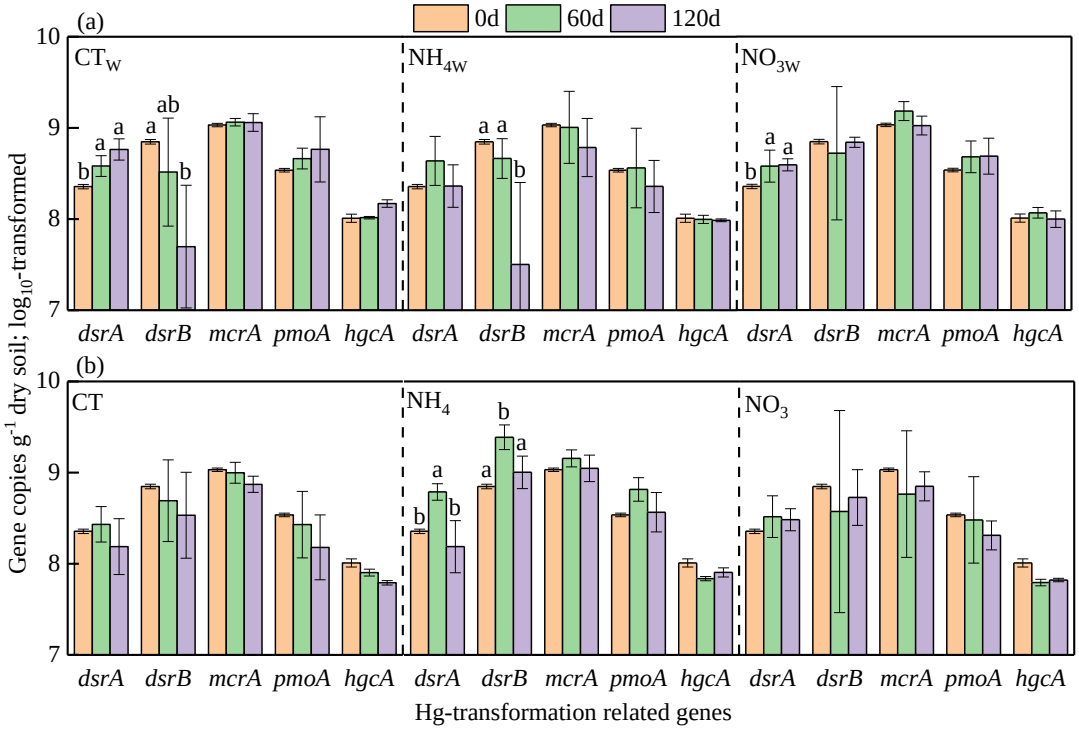


Figure S9. Changes of microbial community composition in soils at 60 d (a, b) and 120 d (c, d). (a) and (c) at the phylum level, (b) and (d) at the genus level.

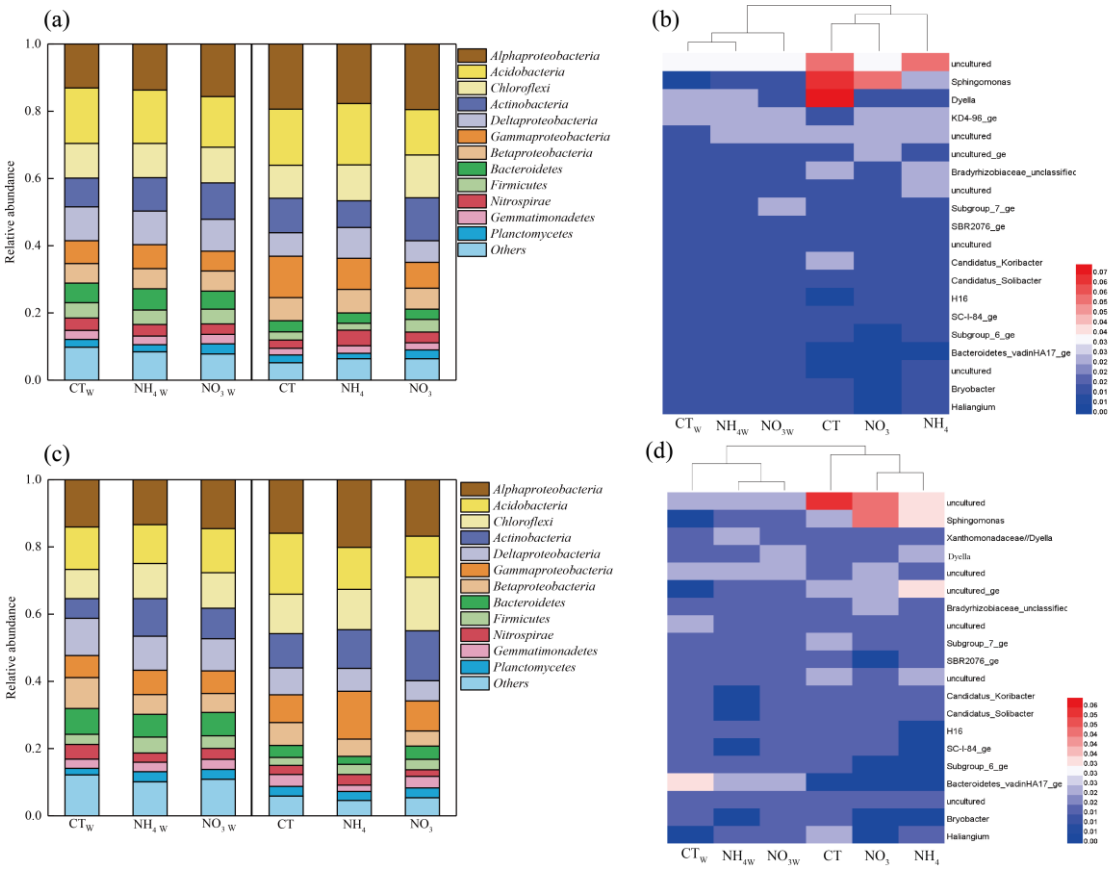


Figure S10. Random forest analysis identifying statistically significant ($p < 0.05$) microbial predictors of MeHg degradation at the family level.

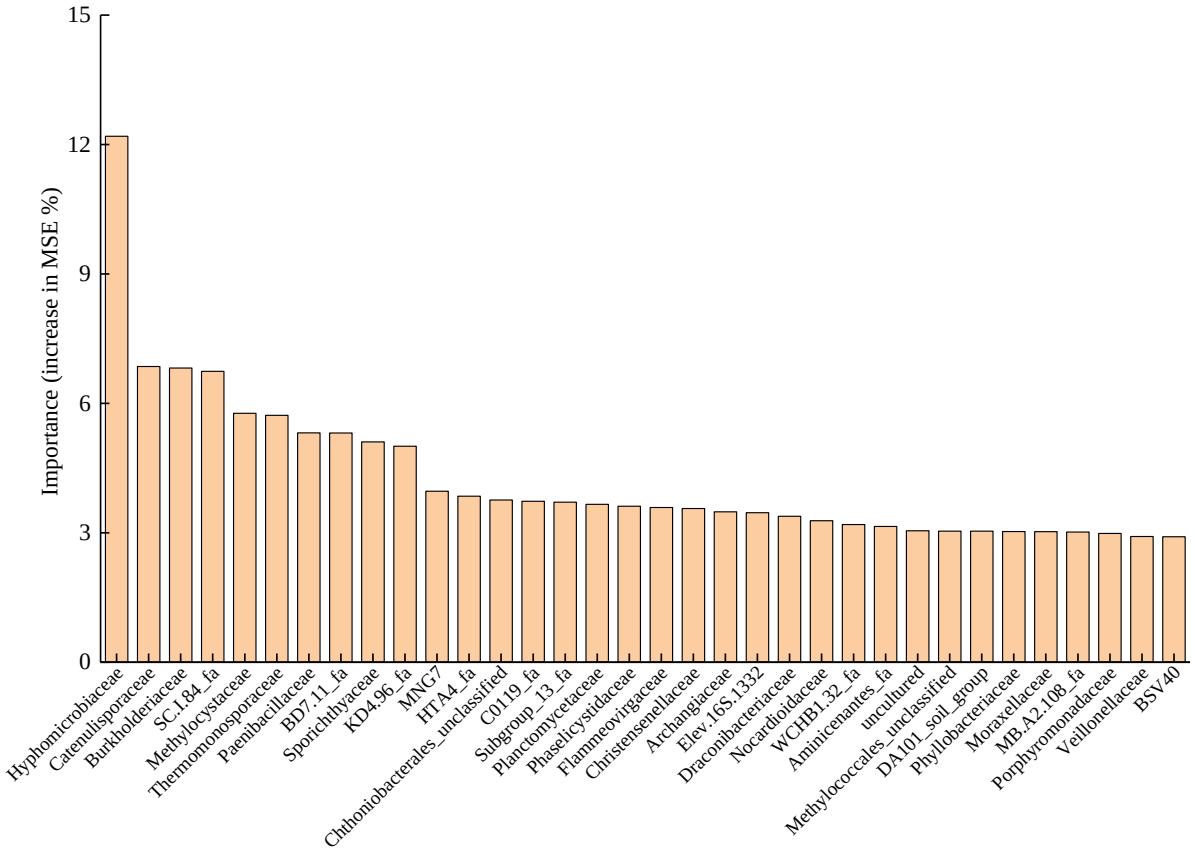


Figure S11. Random forest analysis identifying statistically significant ($p < 0.05$) microbial predictors for MeHg degradation at the family level within the Mod 3.

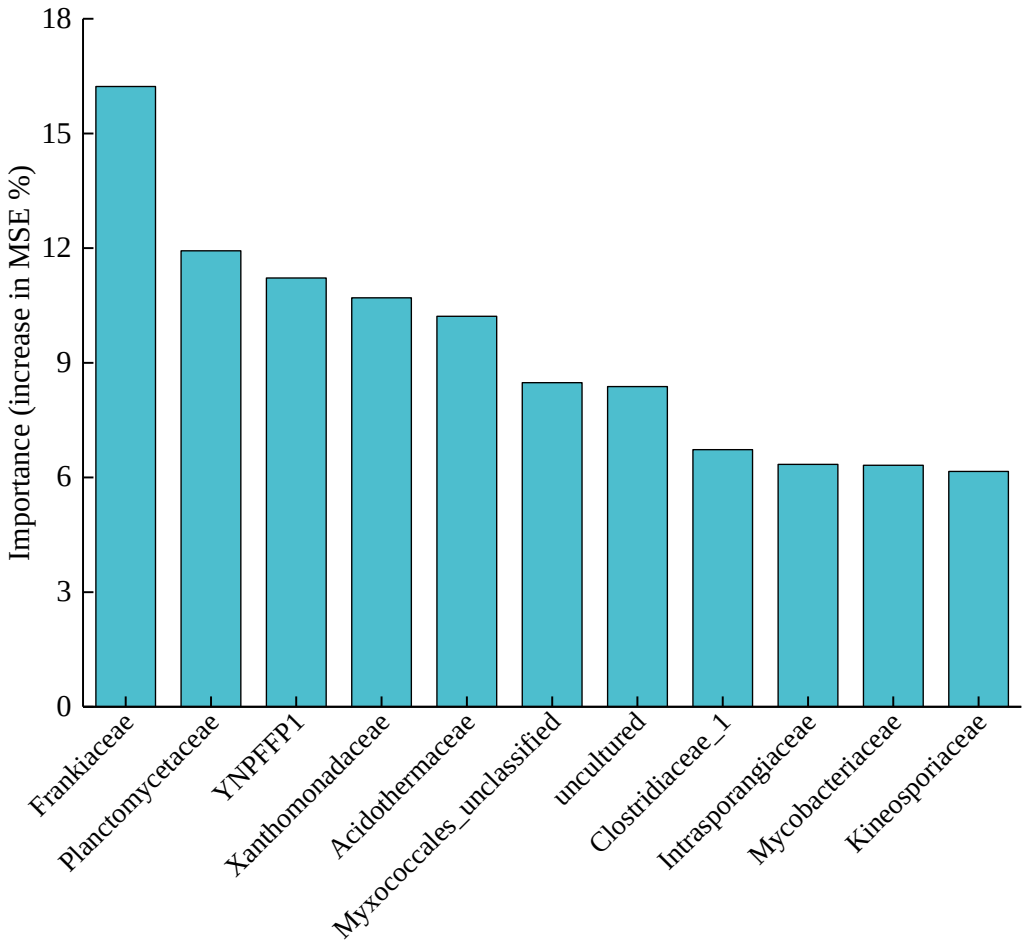


Figure S12. The LEfSe analysis for taxonomic biomarkers at the phylum level. The identified taxa were derived from RNA-based metatranscriptomic analysis.

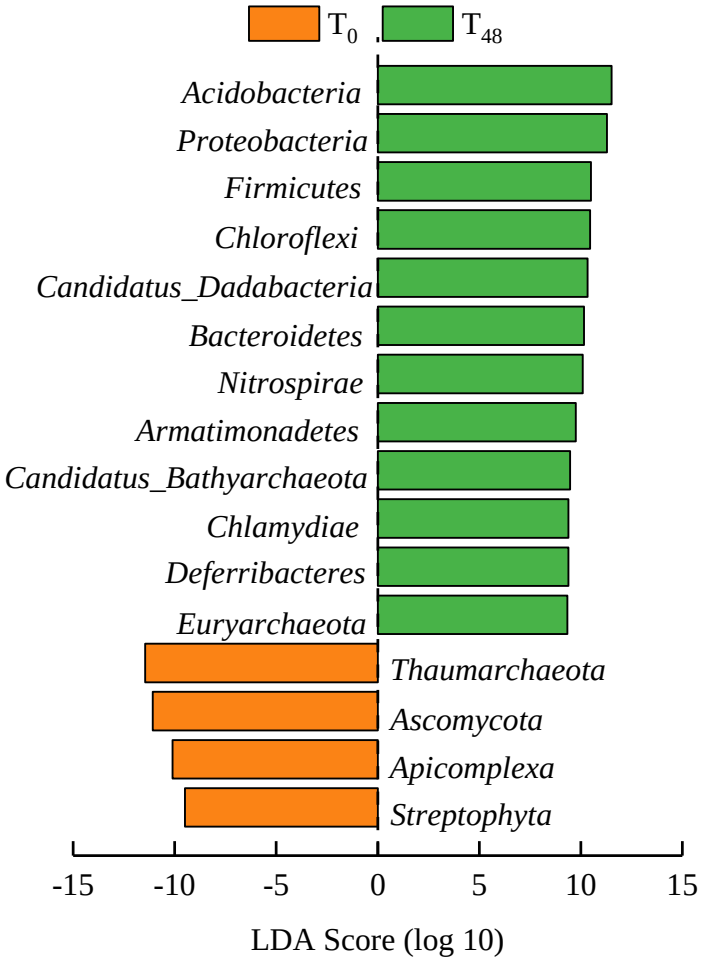
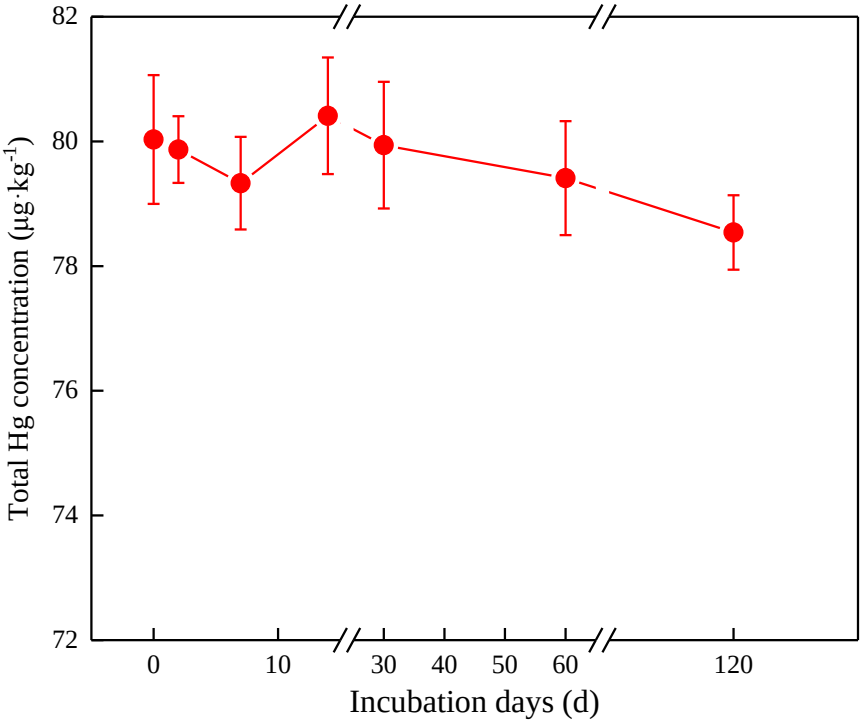


Figure S13. The changes of soil total Hg concentration during 120 d incubation.



Reference

1. Joshi, N.; Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software]. **2011**.

2. Kopylova, E.; Noé, L.; Touzet, H. In *S*ortMeRNA: a new software to filter total RNA for metatranscriptomic or RNA analysis. **2012**.

3. Henschel, R.; Lieber, M.; Wu, L.-S.; Nista, P. M.; Haas, B. J.; LeDuc, R. D. In Trinity RNA-Seq assembler performance optimization. Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond, **2012**; ACM: 2012; p 45.

4. Jiang, Y.; Xiong, X.; Danska, J.; Parkinson, J. Metatranscriptomic analysisof diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* **2016**, *4*, 2.

5. Ismail, W. M.; Ye, Y.; Tang, H. In Gene finding in metatranscriptomic sequences. *BMC Bioinformatics*. **2014**, *15* (9), S8.

6. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012,** *28* (23), 3150-3152.

7. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997,** *25* (17), 3389-3402.

8. Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren MJ, Salzberg, S. L.; Wold, B. J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515.

9. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol*. **2011**, *7* (10), e1002195.

10. Christensen, G. A.; Wymore, A. M.; King, A. J.; Podar, M.; Hurt, R. A.; Santillan, E. U.; Soren, A.; Brandt, C. C.; Brown, S. D.; Palumbo, A. V. Development and validation of broad-range qualitative and clade-specific quantitative molecular probes for assessing mercury methylation in the environment. *Appl. Environ. Microbiol.* **2016,** *82* (19), 6068-6078.

11. Wagner, M.; Roger, A. J.; Flax, J. L.; Brusseau, G. A.; Stahl, D. A. Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* **1998,** *180* (11), 2975-2982.

12. Geets, J.; Borremans, B.; Diels, L.; Springael, D.; Vangronsveld, J.; Van Der Lelie, D.; Vanbroekhoven, K. DsrB gene-based DGGE for community and diversity surveys of sulfate-reducing bacteria. *J. Microbiol. Methods* **2006,** *66* (2), 194-205.

13. Hales, B. A.; Edwards, C.; Ritchie, D. A.; Hall, G.; Pickup, R. W.; Saunders, J. R. Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl. Environ. Microbiol.* **1996,** *62* (2), 668-675.

14. Costello, A. M.; Lidstrom, M. E. Molecular characterization of functional and phylogenetic genes from natural populations of methanotrophs in lake sediments. *Appl. Environ. Microbiol.* **1999,** *65* (11), 5066-5074.

15. Holmes, A. J.; Costello, A.; Lidstrom, M. E.; Murrell, J. C. Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiol. Lett.* **1995,** *132* (3), 203-208.