Environmental Science: Water Research & Technology

Microbial distribution and variation in produced water from separator to storage tank of shale gas wells in Sichuan Basin, China

Supporting Information

Yimeng Zhang [1], Zhisheng Yu [1*], Hongxun Zhang [1], Ian P. Thompson [2]

[1] College of Resources and Environment, University of Chinese Academy of Sciences, 19 A Yuquan Road, Beijing 100049, P.R. China

[2] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

[*]Corresponding Author. Tel: +86 10 88256057; fax: +86 10 88256057; E-mail: yuzs@ucas.ac.cn; College of Resources and Environment, University of Chinese Academy of Sciences, 19 A Yuquan Road, Beijing 100049, P.R. China

This SI includes complete documentation of sequencing data processing, PCR conditions for sequencing, reaction mixture and conditions of qPCR, microbial diversity estimates (Table S1), rarefaction curves of OTUs (Fig. S1), phylogenetic classification for MiSeq Sequencing data at the phylum level (Fig. S2), unweighted pair group method with arithmetic mean (UPGMA) tree of beta diversity based on OTUs (Fig. S3), standard curve of bacteria, archaea, fungi and *dsrA* of qPCR (Fig. S4), and sampling schematic (Fig. S5).

2.3. DNA extraction and sequencing

The PCR reactions were carried out on a GeneAmp 9700 PCR system (Applied Biosystems, Foster City, CA, USA) with the following conditions: 95 ℃ for 2 min; followed by 30 cycles of 95 ℃ for 30 s, 55 ℃ for 30 s, and 72 ℃ for 60 s; and a final extension at 72 ℃ for 5 min.

2.4. Analysis of sequencing data

Initially, low-quality regions (Q<20) were trimmed from the 5' end of sequences and paired ends were merged into contiguous sequences with FLASH in terms of the minimum overlap length of 10 base pair and the maximum error ratio of 0.2. Sequences were demultiplexed, and a further round of quality control was conducted to remove sequences containing ambiguous bases and reads containing low quality bases (Q<25). Non-repetitive sequences were picked from the remaining high-quality sequences along with removing singletons. Operational taxonomic units (OTUs) were defined as clusters of 97% sequence similarity with removing chimeras using Usearch (version 7.1). Taxonomy was assigned using RDP Classifier against Silva database. The resultant OUT abundance tables were rarefied to an even number of sequences per sample to ensure an equal sampling depth.

Bartlett-test in R software was conducted first to confirm that if the data of alpha diversity index (ACE and Shannon) met the requirements of the ANOVA analysis. The results showed that *p* value of Bartlett-tests ranged between 0.162 and 0.992. Therefore, ANOVA analysis could be performed on alpha diversity index.

2.5. Quantitative PCR (qPCR)

Template DNA (1 μL) was used in a reaction mixture of 25 μL containing 12.5 μL of 2× SYBR Green mix (Fermentas, USA), 1 μL of each primer (10 nM), 0.3 μL bovine serum albumin (BSA; 10 mg/mL; New England Biolabs, MA, USA), and 9.2 μL ddH_2O. The PCR was initiated at 95°C for 10 min, followed by 40 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 58°C, and elongation for 30 s at 72°C. Fluorescence signals were collected at 72°C during the elongation step.

Table S1 Diversity estimates of bacteria, archaea and fungi from separators and storage tanks

Station	Sample	Reads	$OTU_{0.03}$	ACE	Chao	Coverage(%)	Shannon	Simpson
Bacteria								
GGS1	GS1	16148	96	115	122	99.86	2.42	0.1426
	WT1	8088	103	111	116	99.84	3.22	0.086
GGS2	GS2	11555	87	156	134	99.73	1.68	0.3395
	WT2	9129	84	93	93	99.87	3.06	0.775
GGS3	GS3	8569	44	55	53	99.88	1.75	0.2749
	WT3	7546	86	105	102	99.75	1.94	0.3648
GGS4	GS4a	9202	90	104	109	99.80	2.73	0.1085
	GS4b	11064	159	175	172	99.78	3.28	0.0669
	WT4	8220	130	145	145	99.76	3.59	0.0476
GGS5	WT5	10728	95	107	104	99.86	2.23	0.2144
GGS6	WT6	7707	84	94	97	99.81	2.88	0.0908
Archaea								
GGS1	GS1	12917	7	8	7	99.99	0.13	0.9555
	WT1	9227	4	0	4	99.99	0.48	0.7135
GGS2	GS2	13871	6	6	6	99.99	0.46	0.7507
	WT2	17784	7	7	7	99.99	0.4	0.8184
GGS3	GS3	12073	10	13	11	99.98	0.38	0.8414
	WT3	13228	5	0	5	99.99	0.39	0.839
GGS4	GS4a	14290	9	9	9	99.99	0.87	0.5464
	GS4b	10371	8	9	8	99.99	0.32	0.8377
	WT4	16161	10	11	11	99.99	0.05	0.9871
GGS5	WT5	17779	4	4	4	99.99	0.26	0.8803
GGS6	WT6	15220	13	13	13	99.99	0.74	0.6346
Fungi								
GGS1	GS1	BDL						
	WT1	19643	19	20	19	99.99	2.13	0.1858
GGS2	GS2	14654	18	19	18	99.99	1.39	0.42
	WT2	14964	23	23	23	99.99	2.12	0.159
GGS3	GS3	BDL						
	WT3	13348	13	13	13	99.99	1.42	0.3935
GGS4	GS4a	15196	20	20	20	99.99	2.13	0.1524
	GS4b	18666	26	27	26	99.99	1.64	0.3679
	WT4	17508	29	29	29	99.99	1.82	0.252
GGS5	WT5	11643	20	21	20	99.99	1.57	0.3583
GGS6	WT6	16324	21	22	21	99.99	1.64	0.2632

[a] Diversity and richness estimates are based on ⩾97% sequence identity. [b] BDL, below detection limit. The abbreviations of the samples are the same as those used in Table 1

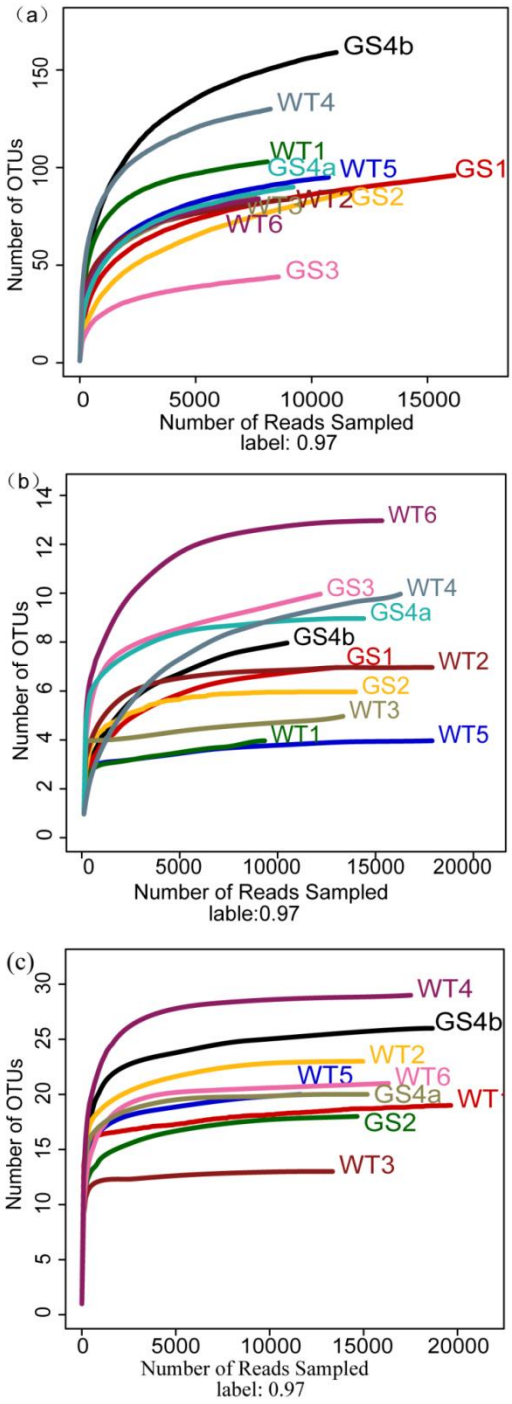
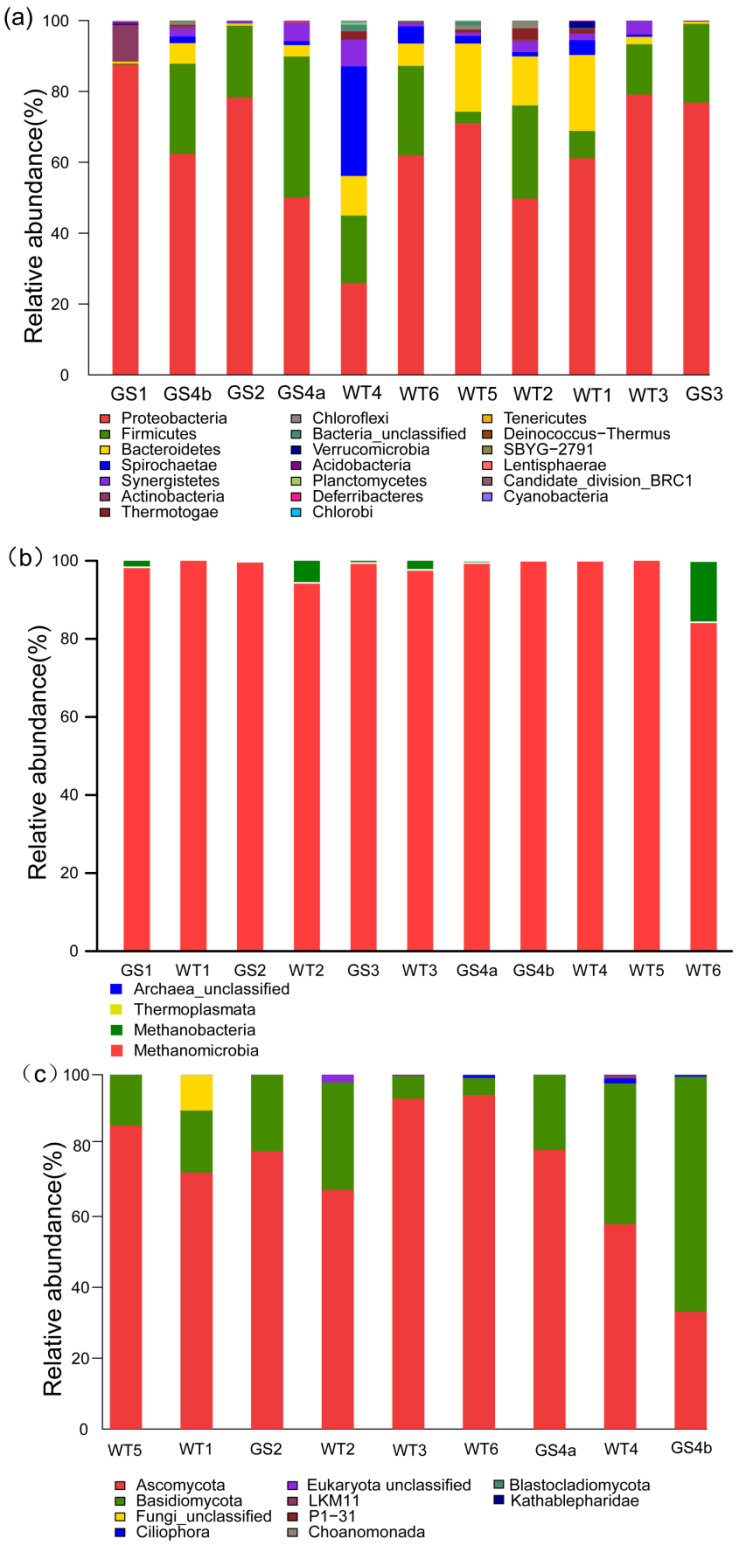


Fig. S1 Rarefaction curves of operational taxonomic units (OTUs) for bacteria (a), archaea (b), and fungi (c) in separator and storage tank samples. The abbreviations of the samples are the same as those used in Fig. 2



72 **Fig. S2** Phylogenetic classification for MiSeq Sequencing of bacteria (a) at the
73 phylum level, archaea (b) at the class level and of fungi (c) at the phylum level
74 obtained from the Silva database project classifier analysis. Fungi were not detected
75 in GS1and GS3. The abbreviations of the samples are the same as those used in Fig. 2

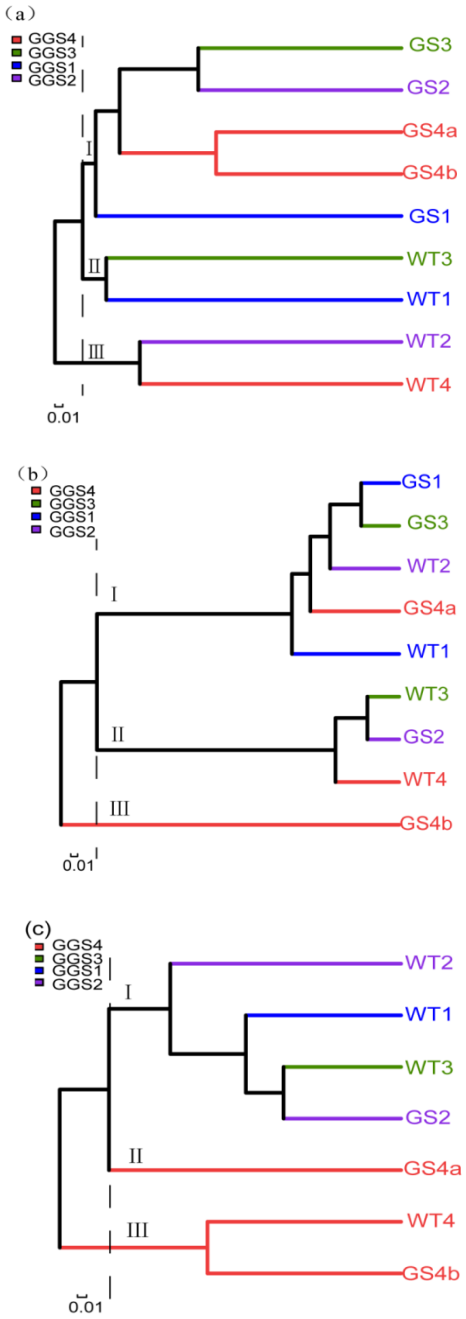


Fig. S3 Unweighted pair group method with arithmetic mean (UPGMA) tree of beta diversity based on operational taxonomic units (OTUs) for bacteria (a), archaea (b), and fungi (c) in the separator and storage tanks. GGS1, GGS2, GGS3 and GGS4 were compared because they all had both separators and storage tanks. Fungi were not detected in GS1 and GS3. The abbreviations of the samples are the same as those used in Fig. 2

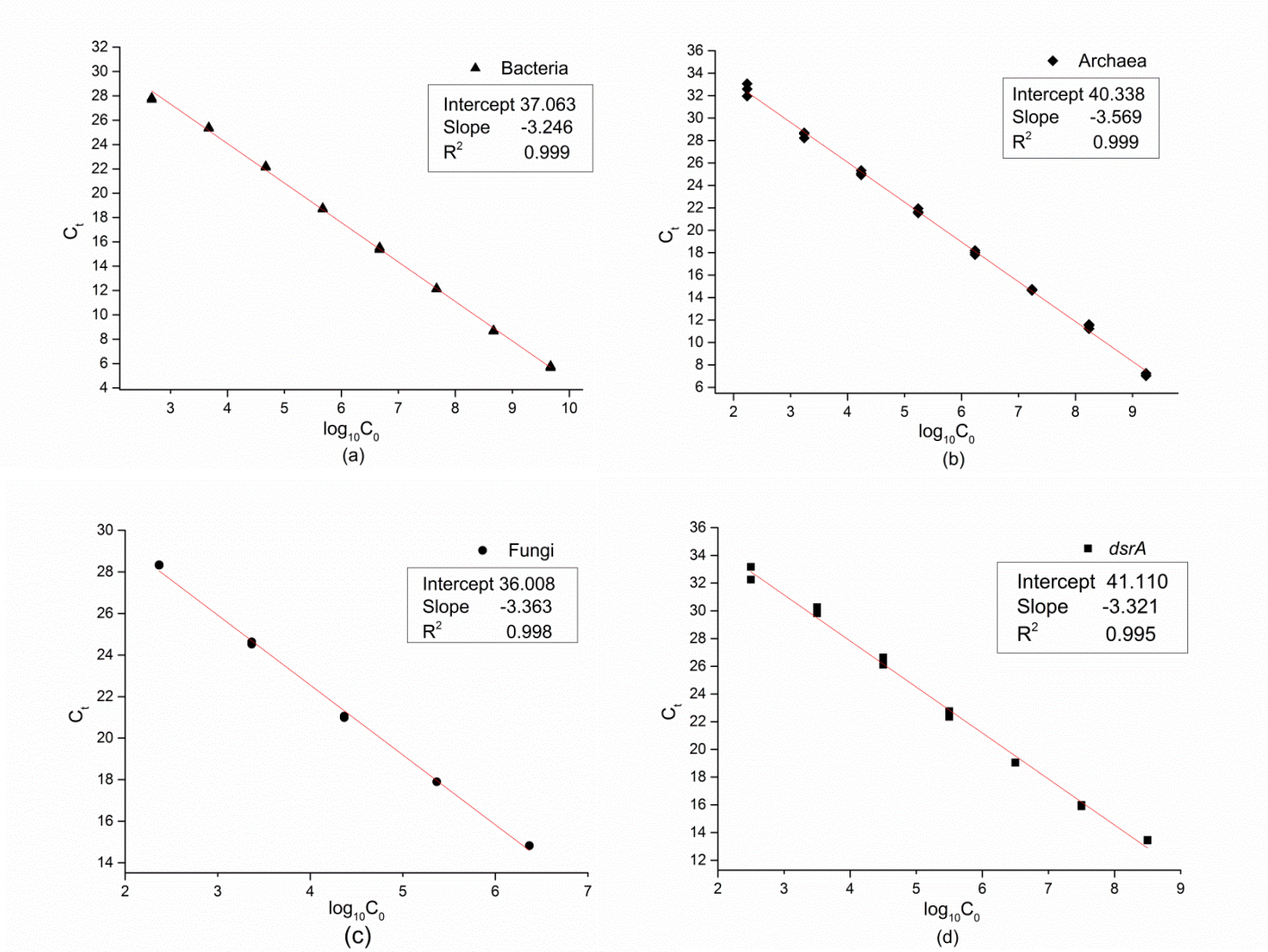
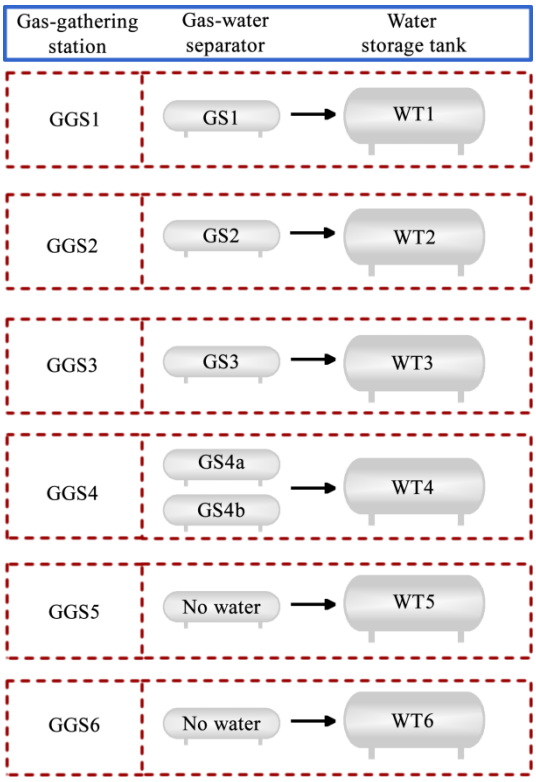


Fig. S4 Standard curve of qPCR for bacteria (a), archaea (b), fungi (c) and *dsrA* genes (d).



107

108 **Fig. S5** Schematic diagram of sampling sites. There was no enough water to be
109 collected in separators at gas-gathering station 5 (GGS5) and 6 (GGS6), thus only
110 water storage tank samples were collected. The abbreviations of the samples are the
111 same as those used in Fig. 2.