

Microbial community and diversity along a salinity gradient of seawater intrusion: the study of microbiological and hydro-geochemical environments in the Pearl River Delta, China

Shilei Sang^{1,4}, Heng Dai^{2,3}, Bill X. Hu^{1*}, Yanru Hao¹, Tong Zhou⁵, Jin Zhang¹*

*corresponding author: heng.dai@jnu.edu.cn (H. D.) and Bill.x.hu@gmail.com (B.X. Hu)

¹ Institute of Groundwater and Earth Sciences, Jinan University, Guangzhou 510632, Guangdong, China

² State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

³ Cele National Station of Observation and Research for Desert-Grassland Ecosystem, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Cele 848300, China

⁴ Department of Ecology, Jinan University, Guangzhou 510632, Guangdong, China

⁵ Laboratory of Marine Organism Taxonomy and Phylogeny, Institute of Oceanology, Chinese Academy of Sciences, 266071, Qingdao, China

Hydrochemical analyses of water samples

The physicochemical parameters such as oxidation-reduction potential (ORP), TDS, pH, and temperature were measured using a portable meter (Hanna Instrument, Milan, Italy). Water samples used for physicochemical analyses were filtered through a sterile 0.45- μm nitrocellulose membrane filter (Millipore, Sigma., Burlington, MA, USA) with a vacuum system.

Cations (K^+ , Ca^{2+} , Na^+ and Mg^{2+}) were measured using Inductively Coupled Plasma Optical Emission Spectrometry (ICAP 7600, ICP-OES; Thermo Fisher Scientific, Waltham, MA, USA). Anions (NO_3^- , Cl^- and SO_4^{2-}) were measured using ion chromatography (Swiss Wantong type 883 chromatograph; Metrohm Schweiz AG, Zofingen, Switzerland). HCO_3^- was measured using acid-based titration analysis (DZ/T 0064.49-93). Total nitrogen (TN) was detected using the alkaline potassium persulfate digestion and UV spectrophotometric method; total phosphorus (TP) was detected using the persulfate digestion and spectrophotometric method; total organic carbon (TOC) was measured using a total carbon analyzer (Elementar, Liquid TOCII; Elementar Analysensysteme GmbH, Langenselbold, Germany)^[1]. Correlation analysis of hydrochemical factors were performed by SPSS ^[2]

Physicochemical analyses were measured according to the research ^[3]. Groundwater samples were collected in 600mL Pyrex Brand square bottles for radiocarbon (^{14}C) and tritium (^3H) analyses. Each bottle was sealed carefully to avoid the existence of bubbles within the water sample. $\delta^2\text{H}$ and $\delta^{18}\text{O}$ were measured using a Finnigan GasBench II Auto carbonate and water device interfaced to a Finnigan MAT DELTAplus XP stable isotope ratio mass spectrometer in the Stable Isotope Laboratory at Florida State University, and reported relative to the VSMOW Standard (Vienna Standard Mean Ocean Water) in permil (‰), with precision ± 2 and $\pm 0.1\%$, respectively. The measurements of ^3H were performed in the Analytical Laboratory of Beijing Research Institute of Uranium Geology and they were electrolytically enriched and measured using the liquid scintillation counting method with an error from ± 0.4 to ± 0.7 TU. The

measurement of ^{14}C was conducted in the Beta Analytic Radiocarbon Lab in Miami (Florida) for the radiocarbon analysis with an error of ± 0.5 pMC.

Molecular analyses for water samples

Total DNA was extracted from 5 L of water filtered through a sterile 0.2- μm nitrocellulose membrane filter (Millipore, Sigma., Burlington, MA, USA) using a vacuum system. A MOBIO PowerSoil® DNA Isolation Kit (Qiagen/MO BIO Laboratories Inc., Carlsbad, CA, USA) was used to extract the DNA. The V4 region of the prokaryotic microbial 16S rRNA gene was amplified by PCR using the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and the reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3')^[4]. The functional gene *mcrA* of the methanogens was amplified by PCR using the forward primer ME1-F (5'-GCMATGCARATHGGWATGTC-3') and the reverse primer ME2-R (5'-TCATKGCRTAGTTDGGRTA-3')^[5]. The functional gene *dsrA* of the sulfate reducers was amplified by PCR using the forward primer dsrA 290-F (5'-CGGCGTTGCGCATTTYCAYACVVT-3') and the reverse primer dsrA 660-R (5'-GCCGGACGATGCAGHTCRTCCTGRWA-3')^{[6][5]58}. PCR reactions were conducted on a BioRad S1000 thermocycler (Bio-Rad Laboratories Inc., Hercules, CA, USA) under the following conditions: 94°C for 5 min; 30 cycles: 94°C for 30 s, 52°C for 30 s and 72°C for 30 s; and 72°C for 10 min. Amplicons were extracted from 1.0% agarose gels and purified using an EZNA Gel Extraction Kit (Omega, Bio-Tek, Norcross, GA, USA) according to the manufacturer's instructions. Libraries were prepared using an NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions, and sequencing was performed on an Illumina HiSeq 2500 system at Magi Gene Technology (Guangzhou, China).

Paired-end raw reads were demultiplexed, quality-filtered by Trimmomatic and merged by the Fast Length Adjustment of SHort reads (FLASH) using the method as described previously ^[7]. Then, the sequences were assigned into operational taxonomic units (OTUs) with a 97% similarity cutoff in the UPARSE

platform ^[8] and chimeric microbial sequences were screened using UCHIME ^[9]. The taxonomy of each 16S rRNA gene sequence was analyzed by the RDP Classifier algorithm (<http://rdp.cme.msu.edu/>) against the Silva 16S rRNA database using a confidence threshold of 70% ^[10].

Rarefaction curves were plotted for each sample to determine the abundance of communities and sequencing data ^[11]. Alpha-diversity analyses including community diversity indexes (Shannon and Simpson), community richness parameters (Chao and ACE), community evenness indexes (Heip) as well as a sequencing depth index (Good's coverage), were calculated using the mothur software ^[12]. The tree figure of groundwater physicochemical characteristics was obtained using the Quantitative Insights Into Microbial Ecology (Qiime 1.7.0) software^[13] based on the Euclidean distance and hierarchical clustering tree of microbial communities on the OTU level based on the Unweighted-Unifrac distance.

In addition, the evolutionary history was concluded among the representative OTUs (abundance > 1% at least one sample) obtained in this study and reference 16S rRNA sequences retrieved from the NCBI GenBank using the Neighbor-Joining method ^[14]. The optimal tree with a sum of branch length = 1.9929 is shown. The phylogenetic tree was drawn to scale using the Maximum Composite Likelihood method ^[15] and the evolutionary distances were in the units of the number of base substitutions per site. The analysis involved 82 nucleotide sequences. Evolutionary analyses were conducted in MEGA 7 ^[16].

Supplementary Table S1 Pearson correlation analysis between major ions of the investigated groundwater samples

Pearson Correlation	Temperature	TOC	TN	TP	NO ₂ ⁻	Cl ⁻	pH	ORP	TDS	NO ₃ ⁻	SO ₄ ²⁻	HCO ₃ ⁻	K ⁺	Na ⁺	Ca ²⁺	Mg ²⁺
Temperature	1	-0.23	0.55	-0.43	-0.32	0.64	0.06	-0.12	0.64	0.18	-0.29	-0.2	0.32	0.55	0.49	0.55
TOC	-0.23	1	0.14	0.84*	-0.26	0.14	0.89*	-0.77	0.14	-0.93**	0.2	-0.6	0.37	0.09	0.09	0.09
TN	0.55	0.14	1	-0.15	-0.09	-0.03	0.37	-0.03	-0.03	-0.15	-0.54	0.26	-0.14	-0.09	-0.31	-0.09
TP	-0.43	0.84*	-0.15	1	0.09	-0.03	0.52	-0.75	-0.03	-0.88	0.46	-0.46	0.32	0.06	0.03	0.06
NO ₂ ⁻	-0.32	-0.26	-0.09	0.09	1	-0.77	-0.6	0.49	-0.77	0.32	-0.37	0.54	-0.77	-0.71	-0.6	-0.71
Cl ⁻	0.64	0.14	-0.03	-0.03	-0.77	1	0.37	-0.6	1**	-0.23	0.37	-0.77	0.89*	0.94**	0.94**	0.94**
pH	0.06	0.89*	0.37	0.52	-0.6	0.37	1	-0.71	0.37	-0.81*	0.09	-0.6	0.49	0.26	0.2	0.26
ORP	-0.12	-0.77	-0.03	-0.75	0.49	-0.6	-0.71	1	-0.6	0.9*	-0.6	0.77	-0.83*	-0.66	-0.54	-0.66
TDS	0.64	0.14	-0.03	-0.03	-0.77	1**	0.37	-0.6	1	-0.23	0.37	-0.77	0.89*	0.94**	0.94**	0.94**
NO ₃ ⁻	0.18	-0.93**	-0.15	-0.88*	0.32	-0.23	-0.81*	0.9*	-0.23	1	-0.46	0.55	-0.55	-0.29	-0.15	-0.29
SO ₄ ²⁻	-0.29	0.2	-0.54	0.46	-0.37	0.37	0.09	-0.6	0.37	-0.46	1	-0.37	0.71	0.6	0.43	0.6
HCO ₃ ⁻	-0.2	-0.6	0.26	-0.46	0.54	-0.77	-0.6	0.77	-0.77	0.55	-0.37	1	-0.77	-0.66	-0.83*	-0.66
K ⁺	0.32	0.37	-0.14	0.32	-0.77	0.89*	0.49	-0.83*	0.89*	-0.55	0.71	-0.77	1	0.94**	0.83*	0.94**
Na ⁺	0.55	0.09	-0.09	0.06	-0.71	0.94**	0.26	-0.66	0.94**	-0.29	0.6	-0.66	0.94**	1	0.89*	1**
Ca ²⁺	0.49	0.09	-0.31	0.03	-0.6	0.94**	0.2	-0.54	0.94**	-0.15	0.43	-0.83*	0.83*	0.89*	1	0.89*
Mg ²⁺	0.55	0.09	-0.09	0.06	-0.71	0.94**	0.26	-0.66	0.94**	-0.29	0.6	-0.66	0.94**	1**	0.89*	1

Significant correlations were marked in **Bold** (* P <0.05, ** P <0.01)

Supplementary Table S2 The relative abundances of dominant phyla (abundance > 1% at least one sample) in all the investigated groundwater samples.

Phylum taxon	Q149	Q146	Q144	Q143	Q130	Q132
<i>Proteobacteria</i>	61.2%	56.1%	60.3%	85.5%	86.9%	84.4%
<i>Firmicutes</i>	37.8%	43.0%	36.3%	12.7%	10.6%	7.4%
<i>Bacteroidetes</i>	0.1%	0.2%	0.8%	0.3%	0.9%	5.8%
<i>Cyanobacteria</i>	0	0	0.3%	0.0%	0.3%	1.2%
others	0.9%	0.7%	2.3%	1.5%	1.3%	1.1%

Supplementary Table S3 The relative abundances of the presentative classes (abundance > 1% at least one sample) in all the investigated groundwater samples.

Class taxon	Q149	Q146	Q144	Q143	Q130	Q132
<i>Gammaproteobacteria</i>	57.7%	53.4%	43.3%	25.1%	42.2%	46.2%
<i>Bacilli</i>	37.8%	43.0%	36.3%	12.7%	10.6%	7.4%
<i>Betaproteobacteria</i>	1.0%	1.0%	4.2%	9.8%	39.9%	29.9%
<i>Epsilonproteobacteria</i>	0.1%	0.1%	0.1%	47.0%	0.1%	1.5%
<i>Alphaproteobacteria</i>	1.9%	1.2%	12.0%	3.2%	4.5%	6.5%
<i>Sphingobacteriia</i>	0.1%	0.1%	0.8%	0.2%	0.8%	5.5%
<i>Cyanobacteria</i>	0	0	0.3%	0.0%	0.3%	1.2%
others	1.3%	1.1%	3.1%	2.0%	1.6%	1.6%

Supplementary Table S4 The relative abundances of the top 50 genera in all the investigated groundwater samples

Genus taxon	Q130	Q132	Q143	Q144	Q146	Q149
<i>Exiguobacterium</i>	10.1%	6.9%	12.3%	34.6%	41.5%	36.6%
<i>Acinetobacter</i>	8.9%	19.6%	5.5%	17.8%	19.4%	16.3%
<i>unclassified_f_Enterobacteriaceae</i>	5.3%	4.0%	6.2%	16.8%	22.7%	20.8%
<i>Sulfuricurvum</i>	0.1%	0.5%	46.7%	0.1%	0.1%	0.1%
<i>Aeromonas</i>	20.3%	17.1%	2.7%	0.3%	0.3%	0.2%
<i>Pseudomonas</i>	2.5%	2.0%	6.7%	7.4%	10.1%	10.4%
<i>unclassified_f_Rhodocyclaceae</i>	13.7%	2.1%	0.5%	2.9%	0.1%	0.1%
<i>norank_f_Gallionellaceae</i>	8.8%	0.6%	2.4%	0.2%	0.2%	0.3%
<i>norank_f_Methylophilaceae</i>	1.1%	8.7%	0.0%	0.0%	0.0%	0.0%
<i>Marinobacter</i>	0.0%	0.0%	0.0%	0.0%	0.0%	8.4%
<i>Acidovorax</i>	6.6%	0.9%	0.1%	0.0%	0.0%	0.0%
<i>Comamonas</i>	2.5%	4.8%	0.1%	0.1%	0.0%	0.0%
<i>norank_f_Rhodocyclaceae</i>	1.9%	3.6%	1.1%	0.1%	0.0%	0.0%
<i>Sediminibacterium</i>	0.2%	5.0%	0.0%	0.6%	0.0%	0.0%
<i>Methylomonas</i>	4.1%	1.4%	0.0%	0.0%	0.0%	0.0%

<i>Methylocystis</i>	0.7%	3.4%	0.0%	1.3%	0.0%	0.0%
<i>unclassified_c__Betaproteobacteria</i>	1.9%	1.8%	0.1%	0.0%	0.0%	0.0%
<i>unclassified_f__Methylophilaceae</i>	0.3%	3.3%	0.0%	0.1%	0.0%	0.1%
<i>norank_f__Hydrogenophilaceae</i>	0.5%	0.1%	2.9%	0.1%	0.0%	0.0%
<i>Bacillus</i>	0.2%	0.2%	0.3%	0.8%	0.9%	0.7%
<i>Magnetovibrio</i>	0.0%	0.0%	0.0%	2.9%	0.0%	0.0%
<i>Thiovirga</i>	0.0%	0.0%	2.2%	0.0%	0.0%	0.0%
<i>Sphingomonas</i>	0.3%	0.5%	0.3%	0.5%	0.3%	0.3%
<i>Thiobacillus</i>	0.1%	0.1%	1.8%	0.1%	0.1%	0.0%
<i>Pseudolabrys</i>	0.0%	0.0%	0.0%	2.0%	0.0%	0.0%
<i>unclassified_f__Rhodospirillaceae</i>	0.1%	0.0%	0.0%	1.8%	0	0.0%
<i>Aquabacterium</i>	0.1%	1.6%	0.0%	0.0%	0.0%	0.0%
<i>unclassified_f__Rhodobacteraceae</i>	0.0%	0.0%	0.6%	0.0%	0.1%	1.0%
<i>Novosphingobium</i>	0.0%	0.1%	0.5%	1.0%	0.0%	0.0%
<i>Hyphomicrobium</i>	1.1%	0.3%	0.0%	0.1%	0.0%	0.0%
<i>norank_o__Obscuribacterales</i>	0.2%	1.2%	0.0%	0.0%	0	0
<i>Sulfurimonas</i>	0.0%	1.1%	0.3%	0.0%	0.0%	0.0%
<i>Vibrio</i>	0.0%	0.0%	0.0%	0.0%	0.1%	1.0%
<i>Enterococcus</i>	0.1%	0.0%	0.1%	0.3%	0.4%	0.3%
<i>norank_f__Nitrosomonadaceae</i>	0.2%	0.3%	0.2%	0.1%	0.1%	0.1%
<i>unclassified_f__Comamonadaceae</i>	0.5%	0.2%	0.2%	0.0%	0.0%	0.0%
<i>norank_o__TRA3-20</i>	0.3%	0.2%	0.1%	0.2%	0.1%	0.1%
<i>Mizugakiibacter</i>	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
<i>Meganema</i>	0.9%	0.0%	0.0%	0.0%	0	0.0%
<i>norank_f__Gemmatimonadaceae</i>	0.1%	0.1%	0.1%	0.2%	0.1%	0.1%
<i>Magnetospirillum</i>	0.0%	0.0%	0.0%	0.7%	0	0
<i>Thioclava</i>	0	0.0%	0.0%	0.1%	0.4%	0.1%
<i>Chryseomicrobium</i>	0.1%	0.2%	0.0%	0.4%	0.1%	0.0%
<i>Methylocaldum</i>	0.1%	0.6%	0	0.0%	0.0%	0.0%
<i>norank_c__Soil_Crenarchaeotic_Group</i>	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%
<i>Bradyrhizobium</i>	0.1%	0.4%	0.0%	0.1%	0.0%	0.0%
<i>unclassified_c__Gammaproteobacteria</i>	0.2%	0.0%	0.3%	0.0%	0.0%	0.0%
<i>Desulfovibrio</i>	0.0%	0.1%	0.0%	0.3%	0.0%	0.2%
<i>norank_f__Anaerolineaceae</i>	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%
<i>Rhodobacter</i>	0.0%	0.5%	0.0%	0.1%	0.0%	0.0%

Supplementary Table S5 The relative abundances of the OTUs (abundance > 1% at least one sample) in all the investigated groundwater samples.

Phylum	Class	Genus	OTU	Q132	Q130	Q143	Q144	Q146	Q149
<i>Firmicutes</i>	<i>Bacilli</i>	<i>Exiguobacterium</i>	OTU388	6.8%	10.1%	12.2%	34.5%	41.3%	36.5%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified <i>Enterobacteriaceae</i>	OTU402	3.9%	5.1%	5.9%	15.5%	21.1%	19.2%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Acinetobacter</i>	OTU348	7.2%	5.8%	5.1%	16.7%	19.1%	15.6%
<i>Proteobacteria</i>	<i>Epsilonproteobacteria</i>	<i>Sulfuricurvum</i>	OTU242	0.2%	0.1%	45.9%	0.1%	0.1%	0.1%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Aeromonas</i>	OTU83	15.8%	19.3%	2.6%	0.2%	0.3%	0.2%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Pseudomonas</i>	OTU395	1.8%	2.3%	3.0%	6.6%	8.9%	8.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	norank <i>Gallionellaceae</i>	OTU47	0.6%	8.8%	2.4%	0.2%	0.2%	0.3%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	norank <i>Methylophilaceae</i>	OTU158	8.7%	1.1%	0.0%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	unclassified <i>Rhodocyclaceae</i>	OTU25	0.2%	6.1%	0.1%	2.7%	0.1%	0.1%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Marinobacter</i>	OTU381	0.0%	0.0%	0.0%	0.0%	0.0%	8.5%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Acinetobacter</i>	OTU59	4.8%	2.4%	0.1%	0.4%	0.1%	0.3%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Acidovorax</i>	OTU162	0.9%	6.6%	0.1%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Comamonas</i>	OTU110	4.5%	1.9%	0.1%	0.0%	0.0%	0.0%
<i>Bacteroidetes</i>	<i>Sphingobacteriia</i>	<i>Sediminibacterium</i>	OTU132	4.7%	0.2%	0.0%	0.7%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	unclassified <i>Rhodocyclaceae</i>	OTU265	1.2%	3.9%	0.3%	0.1%	0.0%	0.1%
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Methylocystis</i>	OTU141	3.4%	0.7%	0.0%	0.9%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Acinetobacter</i>	OTU84	4.1%	0.1%	0.1%	0.1%	0.1%	0.1%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	norank <i>Rhodocyclaceae</i>	OTU189	2.2%	1.1%	0.8%	0.1%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	unclassified <i>Rhodocyclaceae</i>	OTU24	0.5%	3.0%	0.1%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Methylomonas</i>	OTU21	1.3%	2.1%	0.0%	0.0%	0.0%	0.0%

<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	unclassified <i>Betaproteobacteria</i>	OTU36	1.7%	1.6%	0.0%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	unclassified <i>Methylophilaceae</i>	OTU101	2.9%	0.2%	0.0%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Pseudomonas</i>	OTU225	0.1%	0.0%	2.1%	0.2%	0.2%	0.6%
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Magnetovibrio</i>	OTU311	0.0%	0.0%	0.0%	2.9%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Acinetobacter</i>	OTU287	2.0%	0.2%	0.2%	0.4%	0.1%	0.2%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	norank <i>Rhodocyclaceae</i>	OTU86	1.4%	0.8%	0.4%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Acinetobacter</i>	OTU145	1.5%	0.5%	0.1%	0.2%	0.1%	0.1%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Pseudomonas</i>	OTU152	0.1%	0.1%	1.3%	0.2%	0.2%	0.5%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Thiovirga</i>	OTU198	0.0%	0.0%	2.2%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Methylomonas</i>	OTU41	0.2%	2.0%	0.0%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Thiobacillus</i>	OTU244	0.1%	0.1%	1.8%	0.1%	0.1%	0.0%
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Pseudolabrys</i>	OTU106	0.0%	0.0%	0.0%	1.9%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Aquabacterium</i>	OTU155	1.6%	0.1%	0.0%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	unclassified <i>Rhodospirillaceae</i>	OTU284	0.0%	0.0%	0.0%	1.6%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	norank <i>Hydrogenophilaceae</i>	OTU202	0.0%	0.1%	1.4%	0.0%	0.0%	0.0%
<i>Proteobacteria</i>	<i>Epsilonproteobacteria</i>	<i>Sulfurimonas</i>	OTU130	1.1%	0.0%	0.3%	0.0%	0.0%	0.0%
<i>Cyanobacteria</i>	<i>Cyanobacteria</i>	norank <i>Obscuribacterales</i>	OTU98	1.0%	0.2%	0.0%	0.0%	0.0%	0.0%
others	others	others	others	13.6%	13.5%	11.6%	13.6%	7.7%	9.6%

Supplementary Table S6 The relative abundances of the presentative orders (abundance > 1% at least one sample) in all the investigated groundwater samples.

Order taxon	Q149	Q146	Q144	Q143	Q130	Q132
<i>Bacillales</i>	37.38%	42.47%	35.73%	12.60%	10.43%	7.29%
<i>Pseudomonadales</i>	26.69%	29.60%	25.24%	12.27%	11.46%	21.68%
<i>Enterobacteriales</i>	20.78%	22.75%	16.82%	6.24%	5.30%	4.02%
<i>Campylobacterales</i>	0.15%	0.14%	0.12%	46.97%	0.12%	1.52%
<i>Aeromonadales</i>	0.21%	0.33%	0.27%	2.67%	20.31%	17.13%
<i>Rhodocyclales</i>	0.14%	0.17%	3.02%	1.67%	15.64%	5.84%
<i>Burkholderiales</i>	0.17%	0.18%	0.25%	0.51%	10.47%	8.19%
<i>Nitrosomonadales</i>	0.40%	0.39%	0.44%	2.66%	9.20%	1.27%
<i>Methylophilales</i>	0.12%	0.08%	0.08%	0.06%	1.63%	12.15%
<i>Rhizobiales</i>	0.27%	0.22%	4.18%	0.92%	3.47%	4.67%
<i>Alteromonadales</i>	8.52%	0.05%	0.03%	0.02%	0.04%	0.03%
<i>Methylococcales</i>	0.11%	0.12%	0.45%	0.15%	4.40%	2.77%
<i>Sphingobacteriales</i>	0.08%	0.12%	0.79%	0.18%	0.85%	5.52%
<i>Rhodospirillales</i>	0.04%	0.00%	5.68%	0.04%	0.45%	0.14%
<i>Hydrogenophilales</i>	0.05%	0.10%	0.13%	4.64%	0.64%	0.22%
<i>Sphingomonadales</i>	0.47%	0.45%	1.68%	1.40%	0.42%	0.77%
<i>unclassified_Betaproteobacteria</i>	0.03%	0.03%	0.01%	0.05%	1.91%	1.84%
<i>Rhodobacterales</i>	1.11%	0.54%	0.27%	0.77%	0.05%	0.51%
<i>Chromatiales</i>	0.01%	0.00%	0.03%	2.65%	0.01%	0.01%
<i>Obscuribacterales</i>	0.00%	0.00%	0.00%	0.00%	0.23%	1.24%
others	3.28%	2.28%	4.77%	3.50%	2.97%	3.22%

References:

- [1] Zhang Q, Wang X, Hou P, et al: Quality and seasonal variation of rainwater harvested from concrete, asphalt, ceramic tile and green roofs in Chongqing, China, *Journal of environmental management*, 2014, 132: 178-187.
- [2] Beaumont R: An introduction to principal component analysis & factor analysis using SPSS 19 and R (psych package), *Factor Analysis and Principal Component Analysis (PCA)*, 2012, 24(8-9).
- [3] Sang S, Zhang X, Dai H, et al: Diversity and predictive metabolic pathways of the prokaryotic microbial community along a groundwater salinity gradient of the Pearl River Delta, China, *Scientific Reports*, 2018, 8(1): 17317.
- [4] Kozich J J, Westcott S L, Baxter N T, et al: Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform, *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 2013, 79(17): 5112-5120.
- [5] Roesch C, Mergel A, Bothe H: Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil, *Applied and Environmental Microbiology*, 2002, 68(8): 3818-3829.
- [6] Perrault E M, Pereyra L P, Hiibel S, et al: Response and recovery of sulfate-reducing biochemical reactors from aerobic stress events, *Proceedings America Society of Mining and Reclamation*, 2009: 1041-1062.
- [7] Li Y, Chen Y, Yang J, et al: Effects of substratum type on bacterial community structure in biofilms in relation to settlement of plantigrades of the mussel *Mytilus coruscus*, *International Biodeterioration & Biodegradation*, 2014, 96: 41-49.
- [8] Edgar R C: UPARSE: highly accurate OTU sequences from microbial amplicon reads, *NATURE METHODS*, 2013, 10(10): 996.
- [9] Edgar R C, Haas B J, Clemente J C, et al: UCHIME improves sensitivity and speed of chimera detection, *BIOINFORMATICS*, 2011, 27(16): 2194-2200.
- [10] Quast C, Pruesse E, Yilmaz P, et al: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, *NUCLEIC ACIDS RESEARCH*, 2013, 41(D1): D590-D596.
- [11] Amato K R, Yeoman C J, Kent A, et al: Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes, *ISME JOURNAL*, 2013, 7(7): 1344-1353.
- [12] Schloss P D, Westcott S L, Ryabin T, et al: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, *Applied and environmental microbiology*, 2009, 75(23): 7537-7541.
- [13] Caporaso J G, Kuczynski J, Stombaugh J, et al: QIIME allows analysis of high-throughput community sequencing data, *Nature methods*, 2010, 7(5): 335.
- [14] Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees., *Molecular biology and evolution*, 1987, 4(4): 406-425.
- [15] Tamura K, Nei M, Kumar S: Prospects for inferring very large phylogenies by using the neighbor-joining method, *Proceedings of the National Academy of Sciences*, 2004, 101(30): 11030-11035.
- [16] Kumar S, Stecher G, Tamura K: MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, *Molecular biology and evolution*, 2016, 33(7): 1870-1874.