Supporting Information

Bacterial survival strategies in an alkaline tailing site and the physiological mechanisms of dominant phylotypes as revealed by metagenomic analyses

Weimin Sun^{1*†}, Enzong Xiao^{2†}, Max Häggblom³, Valdis Krumins⁴, Yiran Dong⁵, Xiaoxu Sun¹, Fangbai Li¹, Qi

Wang¹, Baoqin Li¹, Bei Yan¹

- Guangdong Key Laboratory of Integrated Agro-environmental Pollution Control and Management, Guangdong Institute of Eco-Environmental Science & Technology, Guangzhou 510650, China
- 2. Key Laboratory of Water Quality and Conservation in the Pearl River Delta, Ministry of Education, School of Environmental Science and Engineering, Guangzhou University, Guangzhou 510006, China
- 3. Department of Biochemistry and Microbiology, Rutgers University, New Brunswick NJ 08901, USA
- 4. Department of Environmental Sciences, Rutgers University, New Brunswick, NJ 08901, USA
- 5. Institute for Genomic Biology, University of Illinois, Urbana-Champaign, Urbana IL 61801, USA

*Correspondence to:

Dr. Weimin Sun

808 Tianyuan Road, Guangzhou, Guangdong, China. Phone: 86-020- 87024633. Fax: 86-020-87024123. E-mail: wmsun@soil.gd.cn

†These authors contributed equally to this work.

The Supporting Information contains detailed materials and methods, 8 tables and 15 figures. The supplementary materials have a total of 33 pages.

Geochemical analysis

Prior to analysis, all pond sediment samples were freeze-dried for 48 h and passed a 200-mesh sieve after thoroughly ground. pH was measured using a calibrated HACH HQ30d pH meter (HACH, Loveland, USA) by mixing 10 g ground samples with 25 ml distilled water. Nitrate and sulfate were measured by ion chromatography (DIONEX ICS-40, Sunnyvale, CA, USA). Total sulfur (TS), soluble sulfur (SS), total carbon (TC), and total nitrogen (TN) in the samples were determined by an elemental analyzer (vario MACRO cube, Elementar, Hanau, Germany) directly, and total organic carbon was determined after removing inorganic carbon with 1M HCl (GR, Kemiou, Tianjing). Details for these measurements were described in our previous study (Sun et al., 2016). Total Fe (Fetot) and Fe(II) were measured by a spectrophotometric method as described in our previous study (Sun et al., 2015a; Sun et al., 2015b). To determine trace elements, tailing pond sediment samples were fully digested with HNO₃ and HF (5:1, v/v) (Edgell, 1989), while the bioavailable fractions of metals were extracted using 1M HNO₃ (Brady et al., 2016). The total trace metal(loid)s and bioavailable metal(loid)s were determined in the extracts by ICP-MS (Agilent, 7700x, California, USA). Certified reference materials (SLRS-5, National Research Council, Canada) and internal standards (Rh, 500 μ g/L) were used for the quality control of ICP-MS (Sun et al., 2016).

Illumina MiSeq sequencing of 16S rRNA genes

Paired-end reads were merged from forward and reverse raw reads using FLASH (Magoč and Salzberg, 2011). The primer sequences were removed by cutadapt (version 1.16) and the paired-end reads were then quality filtered by QIIME (Quantitative Insights Into Microbial Ecology, version 1.7.0) (Caporaso et al., 2010; Bokulich et al., 2013). Chimeric sequences were then removed using UCHIME (Haas et al., 2011). The raw reads were filtered using split_libraries_fastq.py in QIIME (V1.7.0) with the following criteria: the maximum number of consecutive low quality base calls: r=3; the percentage of consecutive high-quality base calls: p=0.75; the minimum Phred quality score; q=20; the maximum number of ambiguous bases (typically coded as N): n=0. Operational taxonomic units (OTUs) were clustered at 97% similarity using UPARSE, and taxonomies were assigned to these OTUs using the RDP classifier (Version 2.2) and searching against the latest version of GreenGenes Database

(DeSantis et al., 2006; Wang et al., 2007). Alpha-diversity indices were calculated by QIIME with a sequence depth of 37,467/sample. Beta-diversity analyses were performed to compare microbial compositional differences across three zones. Both weighted and unweighted UniFrac distance matrices were calculated using the Illumina 16S rRNA gene dataset (Lozupone et al., 2011). Principal Coordinate Analysis (PCoA) and Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering were then performed using unweighted and weighted UniFrac distance matrices to discern microbial compositional differences (Kuczynski et al., 2012).

Numerical analyses

A Random Forest (RF) ensemble model was performed to estimate the relative influence (%) of environmental factors on alpha diversity (Chao1 index) using the R package 'randomForest' (Wang et al., 2015). According to previous studies, RF does not require a pre-selection of variables because it has good predictive performance even when most variables are noise (Wang et al., 2015). Therefore, all measured environmental parameters were used as inputs. The agreement between actual measurements and predictions were verified by the threshold-independent receiver operating characteristic (ROC) to calculate the average area under the curve (Fielding and Bell, 2002). Co-occurrence networks were generated to visualize both nutrient- and metal-microbe interactions as described previously (Sun et al., 2018). The co-occurrence networks were visualized using Gephi as described previously (Newman, 2006a; Newman, 2003). Parameters of the network (e.g. number of nodes and edges, average path length, network diameter) were calculated using igraph in R (Csardi and Nepusz, 2006).

Taxonomic Annotation for the Genomic Bins

All bins were assigned to appropriate taxonomic levels by different methods in the following order: (i) genome bins were translated in silico to protein sequences, which were input in PhyloPhlAn (Segata et al., 2013); (ii) open reading frames (ORFs) predicted from each genome bin were checked by sequence similarity to the NCBI non-redundant (nr) database using BlastX with a maximum allowed e-value of 10-5. The results of blastX were parsed by MEGAN (Huson et al., 2007) to determine the phylogeny of recovered bins. The minimum bit score used for the analysis was 60 and a minimum support of 5% for each taxonomic category was used for the LCA algorithm.

A genome bin was assigned to a taxon when at least 85% of the identified ORFs resulted in a concordant taxonomy. Phylogenetic trees were constructed with recovered genome bins and their taxa genome based on the results of blastX using the FastTree algorithm with default settings by PhyloPhlAn.

Functional gene analysis

Metabolic coding potentials of genome bins were explored by Prodigal (Hyatt et al., 2010). EggNOG-mapper (version 4.5.1) and KOBAS (Kanehisa and Goto, 2000) were used to predict function for the identified ORFs. Gene sequences of putative proteins of interest-those related to nutrient cycling and heavy metal resistance and cycling were recruited from the information of the ORFs and verified by comparing with KEGG and the nr database. The abundances of genes of interest in each bin were determined by Bowtie 2.

Sample ID	Zones	Metagenome Analysis	Name in Metagenome
CSWK01	Zone I	No	
CSWK02	Zone I	No	
CSWK03	Zone I	No	
CSWK04	Zone I	Yes	Zone I (S1)
CSWK05	Zone I	No	
CSWK06	Zone I	No	
CSWK07	Zone I	No	
CSWK08	Zone I	Yes	Zone I (S2) *
CSWK10	Zone I	No	
CSWK11	Zone I	No	
CSWK12	Zone I	No	
CSWK13	Zone I	No	
CSWK14	Zone II	No	
CSWK15	Zone II	No	
CSWK16	Zone II	No	
CSWK17	Zone II	Yes	Zone II
CSWK18	Zone II	No	
CSWK19	Zone II	No	
CSWK20	Zone II	No	
CSWK21	Zone II	No	
CSWK22	Zone II	No	
CSWK23	Zone II	No	
CSWK24	Zone II	No	
CSWK25	Zone II	No	
CSWK26	Zone II	No	
CSWK27	Zone II	No	
CSWK28	Zone III	Yes	Zone III
CSWK29	Zone III	No	
CSWK30	Zone III	No	
CSWK31	Zone III	No	
CSWK32	Zone III	No	
CSWK33	Zone III	No	
CSWK34	Zone III	No	
CSWK35	Zone III	No	

Table S1. Summary of the assignment of the sample name to Zones and metagenome analysis.

*Zone (S2) was selected for binning.

Table S2. Table summarizing the analyzed reads in each step of metagenomic analysis.

Metagenome	Raw Sequence	Deduplicate *	Input Read Pairs	Qualified sequences **
Zone I (S1)	28367000	1800533 (6.35%)	26566467	23802997
Zone I (S2)	31377551	1752905 (5.59%)	29624646	25731711
Zone II	24292225	1151253 (4.74%)	23140972	20020540
Zone III	27015490	1302928 (4.82%)	25712562	22247502

*Duplicates were removed by PRINSEQ (version 0.20.4) ** Low quality reads were removed by Trimmomatic (version 0.36) to obtain qualified sequences

Total Contigs 82,490 Total assembled bases 125,562,425 bp (~125 Mb) Average contig length 1,522 bp 756 bp Median contig length 383,814 bp Max length Min length 500 bp

 Table S3. Table summarizing the results of metagenomic-binning.

Sample	pН	Eh	Sulfate	Fetot	Fe(II)	TN	ТС	TH	TS	TOC	SS
		mV	mg/L	mg/kg	mg/kg	%	%	%	%	%	%
CSWK01	8.5±0.02	125±5	735±42	n.a.	n.a.	0.07 ± 0.02	4.1±0.7	0.3 ± 0.02	1 ± 0.05	0.3 ± 0.01	0.9 ± 0.04
CSWK02	8.7±0.08	127±8	899±27	120±9	111±7	0.06	4.2±0.3	0.2±0.03	1.1 ± 0.7	0.2 ± 0.6	0.2 ± 0.05
CSWK03	4.3±0.3	107±3	617±18	455±41	383±13	0.05 ± 0.01	3.2±0.2	0.2 ± 0.04	2±0.1	0.3 ± 0.02	$0.4{\pm}0.07$
CSWK04	7.7±0.1	92±5	2109±75	927±10	18±5	0.03±0.01	0.01	0.12	2.8±0.9	0.2 ± 0.01	2.2±0.6
CSWK05	8.29±0.09	77±7	2679±85	n.a.	n.a.	0.09 ± 0.05	4.3±0.7	0.3±0.01	3.9±0.8	0.2 ± 0.05	3.1±0.04
CSWK06	8.2±0.7	53±2	1532±98	592±89	18±7	0.26 ± 0.07	3.5±0.1	0.4 ± 0.06	1.9±0.9	0.3 ± 0.02	0.4 ± 0.05
CSWK07	8.09±0.1	47±2	1670±73	n.a.	n.a.	0.06 ± 0.02	2.6±0.7	0.2 ± 0.02	2.4±0.9	0.4 ± 0.05	1.8±0.2
CSWK08	8.4±0.1	43±8	875±67	638±36	13±1	0.04 ± 0.02	2.3±0.7	0.1±0.01	3.1±0.2	0.2 ± 0.02	0.6 ± 0.08
CSWK10	9.4±0.7	13±1	1089±94	n.a.	n.a.	0.06 ± 0.03	5.8±0.6	0.2 ± 0.02	2.0 ± 0.07	2.1±0.3	1.2±0.4
CSWK11	10.8±0.5	-24±1	2006±187	n.a.	n.a.	0.08 ± 0.01	5.8±0.7	0.2±0.01	2.2±0.9	4.5±0.5	1.7±0.2
CSWK12	8.7±0.4	46±1	1347±86	575±37	13±5	0.02 ± 0.01	2.3±0.7	0.08 ± 0.01	3.3±0.6	0.2 ± 0.02	0.9±0.03
CSWK13	8.6±0.6	51±5	414±34	577±28	12±2	0.04 ± 0.01	1.9±0.4	0.1±0.01	3±0.8	0.1±0.01	0.01
CSWK14	8.4±0.04	53±3	1460±25	792±57	15±3	0.04 ± 0.01	2.7±0.8	0.1±0.01	3.1±0.9	0.2±0.01	1.1±0.2
CSWK15	8.4±0.2	60±7	590±18	676±49	13±4	0.05 ± 0.02	3.1±0.5	0.1 ± 0.02	3.4±0.9	0.1±0.01	1.5±0.6
CSWK16	8.4±0.5	62±5	1560±66	1033±30	18±8	0.04 ± 0.03	1.6±0.4	0.1±0.03	3.2±0.4	0.3±0.01	1.2±0.3
CSWK17	8.5±0.5	70±6	190±17	701±20	14±6	0.03 ± 0.01	1.9±0.2	0.09 ± 0.02	2.2±0.5	0.2±0.01	0.01
CSWK18	8.3±0.1	75±2	764±29	635±53	13±3	0.02	2.3±0.7	0.07 ± 0.01	2.5±1.7	0.2±0.03	0.1 ± 0.02
CSWK19	8.3±0.8	77±9	286±39	913±72	16±7	0.03±0.01	1.4±1	0.1±0.04	3.2±0.1	0.1±0.02	0.4 ± 0.06
CSWK20	8.3±0.2	78±7	1040±84	1005±23	17±6	0.04 ± 0.01	1.3±0.1	0.1±001	3.8±0.2	0.4±0.01	1±0.3
CSWK21	8.5±0.6	79±5	1503±89	793±10	13±3	0.04 ± 0.02	1.50	0.1 ± 0.02	3.1±0.7	0.4±0.03	1.5±0.2
CSWK22	8.4±0.1	75±11	76±16	563±23	12±4	0.03	1.4±0.6	0.09 ± 0.01	2.2±0.3	0.3±0.09	0.5±0.2
CSWK23	8.5±0.1	78±9	102±13	609±57	12±4	0.02 ± 0.01	1.5±0.1	0.07 ± 0.01	2±0.09	0.1 ± 0.02	0.4 ± 0.05
CSWK24	8.3±0.2	81±3	241±18	627±85	12±7	0.03 ± 0.02	1.3±0.6	0.1±0.02	2.4±1	0.2 ± 0.01	0.4±0.03
CSWK25	8±0.1	77±8	50±9	573±31	12±2	0.02	1.30	0.08 ± 0.01	2.3±0.7	0.4 ± 0.07	0.2±0.01
CSWK26	8.1±0.8	81±3	1475±79	708±28	16±1	0.03	1.50	0.10	2.1±0.3	0.3±0.04	0.9±0.01
CSWK27	8.2±0.5	81±5	1400±135	714±97	13±3	0.05±0.03	1.6±0.2	0.2 ± 0.04	3.7±0.2	0.2±0.01	1.2±0.3

Table S4. The environmental parameters other than metal-related parameters in 34 samples (mean \pm SD).

CSWK28	8.4±0.3	80±3	104±16	403±27	10±6	0.02	2.5 ± 0.5	0.05 ± 0.03	1.1 ± 0.3	0.1 ± 0.01	0.4 ± 0.03
CSWK29	8.3±0.2	86±5	755±48	481±66	11±1	0.02	2.8±0.1	0.06	3.3±0.1	$0.4{\pm}0.02$	0.5 ± 0.02
CSWK30	8.2±0.2	87±2	1524±79	905±39	19±7	0.03 ± 0.01	2.7±0.2	0.1±0.01	3.1±0.3	0.1 ± 0.01	0.4 ± 0.01
CSWK31	8.2±0.2	89±1	1215±81	693±57	14±2	0.03±0.01	1.7±0.4	0.1±0.01	2.4±0.6	0.2±0.03	0.7 ± 0.04

n.a.: not available

Table S5. The concentrations of major metals and metalloids and their bioavailable fractions in 34 samples (mean \pm SD).

Sample	Mn	Mn _{bio}	Zn	Zn _{bio}	As	As _{bio}	Sb	$\mathrm{Sb}_{\mathrm{bio}}$	Pb	Pb_{bio}	Cu	Cd
	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg	mg/kg
CSWK01	4368±178	2720±85	19635±754	8735±132	3418±443	2335±97	156±60	84±4	2385±281	362±62	360±21	214±2
CSWK02	9251±278	7714±450	41599±853	12573±643	7168±384	2085±152	331±87	226±8	5307±527	313±19	814±43	456±8
CSWK03	7568±178	4906±191	24421±534	23198±618	10642 ± 432	1720±294	317±83	158±12	4938±579	260±38	588±37	263±
CSWK04	10414±259	6762 ± 825	6222 ± 78	3935±156	10779±358	1313±78	766±77	298±14	13399±779	814±31	578±79	55±1
CSWK05	6396±96	3561±278	4028±198	3686±49	7059±296	1273±228	471±21	304±12	8253±605	764±92	351±83	36±2
CSWK06	4223±101	1860 ± 291	22542 ± 148	19783±252	4379±633	1820 ± 30	179±50	65±9	2666±184	299±10	490±29	235±
CSWK07	2867 ± 545	1735±54	15541±672	7698±342	2991±97	1785±90	121±48	n.a.	1687±130	294±16	330±21	165±
CSWK08	6036±328	2413±187	7053±52	4274±133	11602 ± 874	728±29	212±23	69±34	3366±94	150±59	522±38	64±:
CSWK10	2843±78	1258±45	20073±481	16520±199	1512±188	762 ± 88	181±46	125±8	1472±59	107±7	237±33	206±
CSWK11	6289±275	4325±94	42643 ± 488	28973±458	3292±654	1259±44	390±98	191±12	3309±181	194±9	567±55	429±2
CSWK12	5248 ± 372	2503 ± 434	4223±43	1644±49	9540±487	629±31	89±5	25±13	2080±77	105±29	545±61	33±′
CSWK13	3638±69	2543±252	6394±36	2989±97	10844±1991	983±130	131±18	80±9	2557±612	225±45	864±97	54±9
CSWK14	4425±967	4034±235	2063±95	1349±312	5491±797	1444±77	57±9	53±24	2078±66	275±29	26±58	19±2
CSWK15	8130±48	4039±355	3860±195	3475±88	10438 ± 201	1591±76	111±7	65±16	3983±109	272 ± 60	504 ± 60	37±′
CSWK16	2812±229	1758±138	2385±69	1897±76	7592±876	1723±59	129±5	82±27	2060±63	342±99	192±21	19±2
CSWK17	3150±521	2423 ± 550	1650 ± 22	985±25	5325±337	1080±91	89±7	53±40	1553±123	127±8	162±12	13±2
CSWK18	2910±151	2118±195	1099±281	569±84	5396±897	968±51	53±12	40±17	969±16	n.a.	93±6	6±1
CSWK19	3665±164	1329±74	1282±95	872±83	10354±799	963±133	71±17	14±3	1305±96	50±6	134±32	10±0
CSWK20	3307±177	1673±289	987±48	840±68	11073±609	1231±55	167±46	71±2	628±46	50±3	118±12	7±0.
CSWK21	8604±297	2499±431	2069±370	1205±21	11099±851	2065±129	190±19	69±61	2872±713	161±18	167±6	14±2
CSWK22	4435±872	2751±688	1148±96	834±79	7211±286	727±61	117±9	77±9	6737±90	89±9	114±12	8±0.
CSWK23	6701±325	2087±91	2274±88	1011±123	14454±7918	980±101	249±67	102 ± 8	2288±61	95±34	256±77	16±0
CSWK24	3823±78	1856±592	2208±116	1701±358	6826±342	1105±68	95±68	72±3	1396±36	103±5	159±46	13±
CSWK25	7542±91	2189±123	2558±135	873±195	13627±345	566±79	166±13	66±8	2295±95	113±7	245±9	18±0
CSWK26	8375±367	2334±94	2174±68	1201±60	11787±924	915±69	162±7	82±11	2151±79	105±5	208±19	17±3
CSWK27	3521±440	1825±238	1785±185	991±78	8178±548	679±63	129±32	119±19	1403±47	97±4	106±31	10±2
CSWK28	2384±179	1708 ± 451	1587±61	1326±168	3056±476	796±47	294±40	233±68	635±76	50±6	47±5	11±0
CSWK29	3622±58	1179±72	5118±133	939±33	18414±837	316±98	1020 ± 52	411±63	1883 ± 142	63±4	199±20	33±3
CSWK30	2708 ± 246	1020 ± 271	2557±137	618±92	17965±2567	922±57	567±45	120±4	1133±50	35±3	88±12	19±.
CSWK31	5949±89	2109±152	2343±76	1612±670	12393±355	930±130	274±37	153±28	2050±171	125±8	185.±59	18±
CSWK32	2991±289	1633±93	960±25	n.a.	8457±761	1671±41	163±23	123±7	994±82	93±8	59±7	6±0.
CSWK33	2179±492	1346±73	1842 ± 85	n.a.	5960±887	1575±136	396±63	284±19	1564±542	268±38	135±9	14±0
CSWK34	3142 ± 448	1390±145	1387±32	886±71	12072±2061	2151±163	286±10	152±35	1141±49	73±8	83±8	10±0
CSWK35	5362±69	1837±99	2001±58	1172±292	17613±3356	847±64	250±14	118±12	2009±92	119±33	151±6	15±0

n.a.: not available.

Group1 Group2 adonis Beta-dispersion Zone_II 0.181 Zone_III 0.128 Zone_II Zone_I 0.001** 0.688 Zone_III Zone_I 0.001** 0.456

Table S6. Microbial community comparison among three groups using Pair-wise PERMANOVA and Beta dispersion

**p<0.01

#OTU ID	Taxonomy
OTU1	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU2	k_Bacteria; p_[Thermi]; c_Deinococci; o_Thermales; f_Thermaceae; g_Meiothermus; s_
OTU3	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Hydrogenophaga; s_
OTU4	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU5	k_Bacteria; p_Proteobacteria; c_; o_; f_; g_; s_
OTU6	k_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae
OTU7	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Oxalobacteraceae; g_Janthinobacterium
OTU8	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU9	k_Bacteria; p_Chloroflexi; c_Anaerolineae; o_A31; f_S47; g_; s_
OTU10	k_Bacteria; p_Proteobacteria; c_; o_; f_; g_; s_
OTU11	k_Bacteria; p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Micrococcaceae
OTU12	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Erythrobacteraceae
OTU13	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales
OTU14	k_Bacteria; p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Microbacteriaceae; g_Yonghaparkia; s
OTU15	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU16	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU17	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Erythrobacteraceae
OTU18	k_Bacteria; p_Nitrospirae; c_Nitrospira; o_Nitrospirales; f_[Thermodesulfovibrionaceae]; g; s
OTU19	k_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_; f_; g_; s_
OTU20	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria
OTU21	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU22	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Sphingomonadaceae; g_Kaistobacter; s_
OTU23	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Methylophilales; f_; g_; s_
OTU24	k_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae
OTU25	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae
OTU26	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Thiobacterales; f_Thiobacteraceae; g_; s_
OTU27	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Thiobacterales; f_Thiobacteraceae; g_; s_
OTU28	k_Bacteria; p_Chloroflexi; c_Anaerolineae; o_GCA004; f_; g_; s_
OTU29	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU30	k_Bacteria; p_Proteobacteria; c_; o_; f_; g_; s_
OTU31	k_Bacteria; p_Firmicutes; c_Clostridia; o_Natranaerobiales; f_Anaerobrancaceae; g_Dethiobacter; s_
OTU32	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU33	k_Bacteria; p_Bacteroidetes; c_Cytophagia; o_Cytophagales; f_Cyclobacteriaceae; g_; s_
OTU34	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU35	k_Bacteria; p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Geodermatophilaceae

Table S7. Phylogenetic information of top 100 most abundant OTUs for constructing the co-occurrence network.

OTU36	k	Bacteria; p	Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU37	k	Bacteria; p_	_Actinobacteria; cActinobacteria; oActinomycetales; fSporichthyaceae; g; s
OTU38	k	Bacteria; p_	Proteobacteria; c_Alphaproteobacteria; o_Rhodospirillales; f_Acetobacteraceae; g_Roseococcus; s_
OTU39	k	Bacteria; p_	Chloroflexi; cThermomicrobia; oJG30-KF-CM45; f; g; s
OTU40	k	Bacteria; p_	Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU41	k	Bacteria; p_	_Proteobacteria; cAlphaproteobacteria; oCaulobacterales; fCaulobacteraceae; gMycoplana; s
OTU42	k	Bacteria; p_	_Actinobacteria; cThermoleophilia; oSolirubrobacterales; f; g; s
OTU43	k	Bacteria; p_	_Actinobacteria; cThermoleophilia; oGaiellales; fGaiellaceae; g; s
OTU44	k	Bacteria; p_	_Bacteroidetes; cCytophagia; oCytophagales; fCytophagaceae; g; s
OTU45	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oHydrogenophilales; fHydrogenophilaceae; gThiobacillus; s
OTU46	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oHydrogenophilales; fHydrogenophilaceae; gThiobacillus; s
OTU47	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oBurkholderiales; fComamonadaceae
OTU48	k	Bacteria; p_	_Proteobacteria; cAlphaproteobacteria; oRhizobiales; fHyphomicrobiaceae; gRhodoplanes; s
OTU49	k	Bacteria; p_	_Proteobacteria; c; o; f; g; s
OTU50	k	Bacteria; p_	_Proteobacteria; cGammaproteobacteria
OTU51	k	Bacteria; p_	_Chloroflexi; cEllin6529; o; f; g; s
OTU52	k	Bacteria; p_	_Proteobacteria; cAlphaproteobacteria; oSphingomonadales; fSphingomonadaceae; gKaistobacter; s
OTU53	k	Bacteria; p_	_Actinobacteria; cThermoleophilia; oGaiellales; fGaiellaceae; g; s
OTU54	k	Bacteria; p_	Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Rhizobiaceae
OTU55	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oThiobacterales; f; g; s
OTU56	k	Bacteria; p_	_Chloroflexi; cAnaerolineae; oenvOPS12; f; g; s
OTU57	k	Bacteria; p_	_Actinobacteria; cThermoleophilia; oGaiellales; fGaiellaceae; g; s
OTU58	k	Bacteria; p_	_Proteobacteria; cGammaproteobacteria
OTU59	k	Bacteria; p_	_Proteobacteria; cAlphaproteobacteria; oCaulobacterales; fCaulobacteraceae; gMycoplana; s
OTU60	k	Bacteria; p_	_Firmicutes; c_Erysipelotrichi; o_Erysipelotrichales; f_Erysipelotrichaceae; g_PSB-M-3; s_
OTU61	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oHydrogenophilales; fHydrogenophilaceae; gThiobacillus; s
OTU62	k	Bacteria; p_	_Actinobacteria; cActinobacteria; oActinomycetales
OTU63	k	Bacteria; p_	_Nitrospirae; cNitrospira; oNitrospirales; f[Thermodesulfovibrionaceae]; g; s
OTU64	k	Bacteria; p_	_Proteobacteria; cGammaproteobacteria; oEnterobacteriales; fEnterobacteriaceae
OTU65	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oHydrogenophilales; fHydrogenophilaceae; gThiobacillus; s
OTU66	k	Bacteria; p_	_Proteobacteria; cGammaproteobacteria; oXanthomonadales; fXanthomonadaceae
OTU67	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oHydrogenophilales; fHydrogenophilaceae; gThiobacillus; s
OTU68	k	Bacteria; p_	_Chloroflexi; cGitt-GS-136; o; f; g; s
OTU69	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oHydrogenophilales; fHydrogenophilaceae; gThiobacillus; s
OTU70	k	Bacteria; p_	_Proteobacteria; cBetaproteobacteria; oBurkholderiales; fComamonadaceae
OTU71	k	Bacteria; p_	_Proteobacteria; cGammaproteobacteria; o; f; g; s
OTU72	k	Bacteria; p_	_Proteobacteria; cAlphaproteobacteria; oCaulobacterales; fCaulobacteraceae; gPhenylobacterium; s

OTU73	k_Bacteria; p_Proteobacteria; c_Deltaproteobacteria; o_Desulfobacterales; f_Desulfobulbaceae
OTU74	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Sphingomonadaceae
OTU75	k_Bacteria; p_Proteobacteria; c_; o_; f_; g_; s_
OTU76	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria
OTU77	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae
OTU78	k_Bacteria; p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Nocardioidaceae; g_; s_
OTU79	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU80	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_; g_; s_
OTU81	k_Bacteria;
01001	p_Actinobacteria
OTU82	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Rhodocyclales; f_Rhodocyclaceae
OTU83	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria
OTU84	k_Bacteria; p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Nocardioidaceae; g_; s_
OTU85	k_Bacteria; p_Actinobacteria; c_Thermoleophilia; o_Gaiellales; f_Gaiellaceae; g_; s_
OTU86	k_Bacteria; p_Actinobacteria; c_Thermoleophilia; o_Gaiellales; f_Gaiellaceae; g_; s_
OTU87	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Sphingomonadaceae; g_Sphingomonas; s_
OTU88	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Hyphomicrobiaceae; g_Devosia; s_
OTU89	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU90	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae
OTU01	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Sphingomonadaceae; g_Blastomonas;
01091	snatatoria
OTU92	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU93	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Comamonadaceae; g_Hydrogenophaga; s_
OTU94	k_Bacteria; p_Proteobacteria; c_Gammaproteobacteria; o_Legionellales; f_Coxiellaceae; g_Aquicella; s_
OTU95	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Burkholderiaceae
OTU96	k_Bacteria; p_Bacteroidetes; c_Sphingobacteriia; o_Sphingobacteriales; f_Sphingobacteriaceae; g_; s_
OTU97	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Burkholderiales; f_Oxalobacteraceae; g_Janthinobacterium; s_
OTU98	k_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Bradyrhizobiaceae; g_Bradyrhizobium; s_
OTU99	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s_
OTU100	k_Bacteria; p_Proteobacteria; c_Betaproteobacteria; o_Hydrogenophilales; f_Hydrogenophilaceae; g_Thiobacillus; s

Table S8. Characteristics of curat	ed genome bins reconstructed from	n the metagenomes of Zone I	(S2).
------------------------------------	-----------------------------------	-----------------------------	-------

ID	Estimated genome size (bp)	NO. contigs	GC%	N50ª	Completeness% ^b	Suggested phylogeny (NCBI-nr) ^c
Bin5	4,464,105	275	44.08	28631	99.03	Algoriphagus sp.
Bin14	2,363,373	236	59.23	15256	98.31	Meiothermus sp.
Bin11	2,703,486	107	43.16	58849	98.13	Nitrospirae
Bin30	7,138,329	794	64.63	12463	97.81	Betaproteobacteria
Bin6	4,199,700	318	66.35	27517	97.33	Rhodobacteraceae
Bin12	5,490,247	596	69.45	11453	96.3	Xanthomonadaceae
Bin20	9,105,763	615	61.41	16923	94.83	Thiobacillus sp.
Bin34	2,201,304	517	66.9	6144	91.78	Comamonadaceae
Bin17	5,527,712	1792	64.91	3271	83.15	Thiobacillus sp.
Bin2	3,275,629	756	40.5	5585	78.17	Bacteroidetes
Bin21	5,170,673	2897	67.15	1843	74.15	Proteobacteria
Bin18	2,802,239	1703	64.45	1696	71.14	Thiobacillus sp.
Bin27	1,807,108	560	36.98	4119	69.36	Firmicutes
Bin16	2,434,926	406	69.49	8136	67.15	Yonghaparkia sp.
Bin25	2,764,688	1634	70.09	1697	57.64	Proteobacteria
Bin22	4,538,044	846	62.38	5849	57.35	Thiobacillus sp.

^a N50: number of the largest contigs that sum up to 50% of the total sum of bases.

^b Genome bins completeness and contamination were calculated by CheckM.

^c A taxon was assigned when at least 85% of the identified ORFs resulted in a concordant taxonomy of NCBI-nr database.



Figure S1. Location of the tailing pond and the locations of 34 sampling sites in the tailing pond (created by Dr. Enzong Xiao)



Figure S2. The distribution of environmental parameters other than metal(loid)s in three zones of the tailing ponds.



Figure S3. The distance matrix which was used for beta diversity analysis.



Figure S4. Distribution of phyla identified using RDP classifier of bacterial 16S rRNA genes from different mine tailings samples.



Figure S5. Distribution of top 20 most abundant genera in three zones of a mine tailings pond.



Figure S6. Co-occurrence network of nutrient-microbe (A) and metal-microbe (B) interactions based on 16S rRNA gene amplicon sequencing. Only strong and significant nutrient-microbe and metal-microbe interactions are shown in the network (microbe-microbe interactions have been removed in this figure to simplify these networks). Positive correlations are displayed in red and negative correlations are displayed in green.



Figure S7. Percentage of reads annotated to carbon fixation pathway genes for four shotgun metagenomics sequencing libraries (A). Taxonomy of reads annotated to carbon fixation pathway genes for four shotgun metagenomics sequencing libraries at the genus level (B).



Figure S8. Percentage of reads annotated to N cycling pathway genes for four shotgun metagenomics sequencing libraries (A). Taxonomy of reads annotated to N cycling pathway genes for four shotgun metagenomics sequencing libraries at the genus level (B).



Figure S9. Percentage of reads annotated to S cycling pathway genes for four shotgun metagenomics sequencing libraries (A). Taxonomy of reads annotated to S cycling pathway genes for four shotgun metagenomics sequencing libraries at the genus level (B).



Figure S10. Phylogenetic affiliations of the genome bins and their read depths. Neighbor-joining phylogenetic tree of the bins (marked as " \star ") and their closest species based on blastx (" \star "), constructed using deduced protein sequences by approximate-maximum-likelihood method in PhyloPhlAn software.



Figure S11. Abundance of genes relating to carbon fixation and H cycling recovered from binned draft genomes from metagenomes of Zone I (S2).



Figure S12. Abundance of genes relating to N cycling recovered from binned draft genomes from metagenomes of Zone I (S2).



Figure S13. Abundance of genes relating to S cycling recovered from binned draft genomes from metagenomes of Zone I (S2).



Figure S14. Abundance of genes relating to metal resistance recovered from binned draft genomes from metagenomes of Zone I (S2).



Figure S15. Heatmap showing the relative abundances of various carbon fixation pathways in each bin. **References**

- 1. Sun, W.; Xiao, E.; Dong, Y.; Tang, S.; Krumins, V.; Ning, Z.; Sun, M.; Zhao, Y.; Wu, S.; Xiao, T., Profiling microbial community in a watershed heavily contaminated by an active antimony (Sb) mine in Southwest China. Sci Tot Environ 2016, 550, 297-308.
- 2. Sun, M.; Xiao, T.; Ning, Z.; Xiao, E.; Sun, W., Microbial community analysis in rice paddy soils irrigated by acid mine drainage contaminated water. Appl Microbiol Biotechnol 2015, 99 (6), 2911-2922.
- 3. Sun, W.; Xiao, T.; Sun, M.; Dong, Y.; Ning, Z.; Xiao, E.; Tang, S.; Li, J., Diversity of the sediment microbial community in the Aha watershed (Southwest China) in response to acid mine drainage pollution gradients. Appl Environ Microbiol 2015, 81 (15), 4874-4884.
- 4. Edgell, K., USEPA Method Study 37 SW-846 Method 3050 Acid Digestion of Sediments, Sludges, and Soils. US Environmental Protection Agency, Environmental Monitoring Systems Laboratory: 1989.
- 5. Brady, J. P.; Kinaev, I.; Goonetilleke, A.; Ayoko, G. A., Comparison of partial extraction reagents for assessing potential bioavailability of heavy metals in sediments. Marine Pollut Bull 2016, 106 (1–2), 329-334.
- 6. Magoč, T.; Salzberg, S. L., FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatic 2011, 27 (21), 2957-2963.
- 7. Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Pena, A. G.; Goodrich, J. K.; Gordon, J. I., QIIME allows analysis of high-throughput community sequencing data. Nat Method 2010, 7 (5), 335-336.
- 8. Bokulich, N. A.; Subramanian, S.; Faith, J. J.; Gevers, D.; Gordon, J. I.; Knight, R.; Mills, D. A.; Caporaso, J. G., Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Method 2013, 10 (1), 57.
- 9. Haas, B. J.; Gevers, D.; Earl, A. M.; Feldgarden, M.; Ward, D. V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S. K.; Sodergren, E., Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res 2011, 21 (3), 494-504.
- 10. DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L., Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006, 72 (7), 5069-5072.
- 11. Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R., Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007, 73 (16), 5261-5267.
- 12. Lozupone, C.; Lladser, M. E.; Knights, D.; Stombaugh, J.; Knight, R., UniFrac: an effective distance metric for microbial community comparison. ISME J 2011, 5 (2), 169.
- 13. Kuczynski, J.; Stombaugh, J.; Walters, W. A.; González, A.; Caporaso, J. G.; Knight, R., Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protocol Microbiol 2012, 1E. 5.1-1E. 5.20.
- 14. Wang, Q.; Xie, Z.; Li, F., Using ensemble models to identify and apportion heavy metal pollution sources in agricultural soils on a local scale. Environ Pollut 2015, 206, 227-235.
- 15. Fielding, A. H.; Bell, J. F., A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 2002, 24 (1), 38-49.
- 16. Sun, W.; Xiao, E.; Pu, Z.; Krumins, V.; Dong, Y.; Li, B.; Hu, M., Paddy soil microbial communities driven by environment- and microbe-microbe interactions: A case study of elevation-resolved microbial communities in a rice terrace. Sci Tot Environ 2018, 612, 884-893.
- 17. Newman, M. E., The structure and function of complex networks. SIAM Rev 2003, 45 (2), 167-256.
- 18. Newman, M. E., Modularity and community structure in networks. PNAS 2006, 103 (23), 8577-8582.
- 19. Csardi, G.; Nepusz, T., The Igraph Software Package for Complex Network Research. Inter J Complex Sys 2006, complex systems.
- 20. Segata, N.; Börnigen, D.; Morgan, X. C.; Huttenhower, C., PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 2013, 4, 2304.
- 21. Huson, D. H.; Auch, A. F.; Qi, J.; Schuster, S. C., MEGAN analysis of metagenomic data. Genome Res 2007, 17 (3), 377-386.
- 22. Hyatt, D.; Chen, G. L.; Locascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J., Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010, 11 (1), 119.

23. Kanehisa, M.; Goto, S., KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000, 28 (1), 27-30.