

**GeoChip-based analysis of microbial functional gene diversity in a landfill
leachate-contaminated aquifer**

Zhenmei Lu,^{1,2} Zhili He,^{2,4} Victoria A. Parisi,^{3,4} Sanghoon Kang,² Ye Deng,² Joy D.
Van Nostrand,² Jason R. Masoner,⁵ Isabelle M. Cozzarelli,⁶ Joseph M. Suflita,^{3,4} and
Jizhong Zhou^{2,4,7,8,*}

College of Life Sciences, Zhejiang University, Hangzhou 310058, China¹; Institute for
Environmental Genomics²; Institute for Energy and the Environment³; Department of
Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA⁴; US
Geological Survey, Oklahoma City, OK 73116, USA⁵; US Geological Survey, 431
National Center, Reston, VA 20192, USA⁶; and Earth Sciences Division, Lawrence
Berkeley National Laboratory, Berkeley, CA 94720⁷, State Key Joint Laboratory of
Environment Simulation and Pollution Control, School of Environment, Tsinghua
University, Beijing 100084, China⁸

*Corresponding author. Institute for Environmental Genomics, University of
Oklahoma, Norman, OK 73019, Phone (405) 325-6073, Fax (405) 325-7552,
jzhou@ou.edu

Summary of supporting information:

A. MATERIALS AND METHODS S3-S5

B. SUPPORTING RESULTS S6-S7

C. SUPPORTING TABLES S8-S9

1. **Table S1.** Other part of geochemistry variables measured in groundwater from each monitoring well.
2. **Table S2.** Distribution of functional genes detected for major functional gene categories detected by GeoChip 3.0.

D. SUPPORTING FIGURES S10-S15

1. **Figure S1.** Location of the Norman Landfill and sampling wells.
2. **Figure S2.** The normalized signal intensity of genes involved in recalcitrant carbon degradation (A) and labile carbon degradation (B).
3. **Figure S3.** The normalized signal intensity of genes involved in methane production and oxidation.
4. **Figure S4.** The normalized signal intensity of the major detected gene families involved in the N cycling for the monitoring wells.
5. **Figure S5.** Hierarchical cluster analysis of *dsrA* gene, encoding alpha subunit of dissimilatory sulfite reductase.
6. **Figure S6.** The normalized signal intensity of genes involved in the S cycling for the monitoring wells.

E. SUPPORTING REFERENCES S16

A. MATERIALS AND METHODS

Site description. The study site is a closed municipal solid waste landfill on the Canadian River alluvial aquifer in Norman, OK, USA (Cozzarelli *et al.*, 2000; 2011). The landfill, which is unlined and has no leachate collection system, started receiving solid wastes from 1922 until 1985 and was covered with clay cap and vegetated when it was closed. Leachate from the unlined landfill has resulted in a groundwater plume that extends downgradient approximately 250 m from the landfill toward the Canadian River (approximately 15 m/year) and flows directly beneath the wetland (Scholl and Christenson, 1998; Grossman *et al.*, 2002). This site has been biogeochemically characterized as described previously (Cozzarelli *et al.*, 2000; Lorah *et al.*, 2009).

DNA extraction, amplification and labeling. High molecular weight DNA was extracted from the filters by freeze-grinding methods (Zhou *et al.*, 1996) and dissolved in 25 μ l of water. The DNA was quantified with the NanoDrop spectrophotometer (ND-1000, Nanodrop Inc., DE, USA) and Quant-It PicoGreen (invitrogen, Carlsbd, CA). Approximately 100ng of DNA were amplified using the Templphi kit (GE Healthcare, Piscataway, NJ, USA) with the following modifications. Spermidine (0.1mM) and single-stranded binding protein (267ng μ l⁻¹) were added to improve the amplification efficiency (Wu *et al.*, 2006). 3.0 μ g amplified DNA was labeled with a cyanine-5 fluorescent dye (GE Healthcare) using a random priming method (Rhee *et al.*, 2004; Tiquia *et al.*, 2004). Labeled DNA was purified using QIA Quick Purification kit (Qiagen, Valencia, CA, USA) according to

the manufacturer's instructions, measured on NanoDrop spectrophotometer and then dried down in a SpeedVac (ThermoSavant, Milford, MA, USA) at 45°C for 1 h.

GeoChip hybridization, scanning and image analysis. GeoChip 3.0 was used for DNA hybridization analysis of all eight samples. GeoChip 3.0 contains ~27,000 probes covering approximately 57,000 gene variants from 292 functional gene families involved in biogeochemical cycling of carbon (C), nitrogen (N), phosphorus (P) and sulfur (S), energy metabolism, and metal resistance and organic contaminants. All hybridizations were carried out in triplicate on different modules at 45°C for 10 h with 50% formamide using a TECAN HS4800 (US TECAN, Durham, NC, USA) after a 45-min prehybridization with a prehybridization solution (5×SSC, 0.1% SDS and 0.1% bovine serum albumin) as described previously (Wu *et al.*, 2006; Xu *et al.*, 2010). GeoChips were scanned by a ScanArray 5000 Microarray Analysis system (Perkin-Elmer, Wellesley, MA, USA) at 95% laser power and 70% photomultiplier tube gain (PMT). ImaGene version 6.0 (Biodiscovery Inc., El Segundo, USA) was then used for image quantification as described previously (Wu *et al.*, 2006). Intensities of three replicates were normalized with mean signal intensities as described previously (Wu *et al.*, 2008). Spots with a signal-to-noise ratio [SNR = (signal intensity-background intensity)/standard deviation of the background] greater than 2 were used for further analysis.

The triplicate hybridizations for each DNA sample were performed to account for variations generated during amplification, labeling, hybridization, and image analysis. For each gene detected, the following criteria were used to pre-process raw data: (i) a

minimum of two out of three spots; (ii) a maximum ratio of 3.0 for pairwise comparison of all three positive spots; and (iii) an average signal intensity of positive spots (two or three) as the final signal intensity.

B. SUPPORTING RESULTS

Genes suggesting metal resistance. Arsenic (As), barium (Ba), cadmium (Ca), chromium (Cr), cobalt (Co), nickel (Ni), and strontium (Sr) had substantially higher concentrations in wells downgradient from the landfill than in background wells. Metal reduction, especially iron reduction is considered a major functional process in leachate contaminated aquifers (Cozzarelli *et al.*, 2000, 2011), and many genes involved in resistance and/or reduction were detected with high abundance. A total of 109 to 406 genes involved in As, Cd, Cr, Cu, Hg, Ag, Te and Zn resistance were detected in each well along the transect. Among them, there were 11 *copA* genes encoding copper-translocating P-type ATPase for Cu resistance, nine *ChrA* genes encoding chromate transporter, four *czcD* genes encoding cation efflux transporter for Cu, Co and Zn resistance, three *mer* genes encoding mercuric reductase for Hg resistance, three *ZntA* genes encoding heavy metal translocating P-type ATPase for Zn resistance, two *CadA* genes encoding cadmium transporting P-type ATPase for Cd resistance, two *czcA* genes encoding heavy metal efflux pump, one *tehB* gene encoding putative tellurite resistance protein, and one *terZ* gene encoding uncharacterized proteins involved in stress response across all samples.

Genes suggesting nitrogen cycling. Previous studies on nitrogen cycling and transport in the landfill-leachate affected aquifer have indicated that NH_4^+ is the dominant nitrogen species in this system and represents a contaminant likely to be of long-term concern (Cozzarelli *et al.*, 2011; Lorah *et al.*, 2009). The normalized signal intensity of the major gene families involved in N-cycling reveal that most of the

genes (59%-64%) are most closely associated with uncultured microorganisms. Those included, two *nifH* genes encoding a nitrogenase and two *nirK* genes encoding a nitrite reductase responsible for N₂ fixation and denitrification, respectively. A gene encoding a cytochrome c nitrite reductase (*nrfA*) most closely related to the protein responsible for dissimilatory N reduction to ammonium was also detected as were two *ureC* genes encoding a urease responsible for ammonification. Generally, the N-cycling genes, especially N₂ fixation (*nifH*) and denitrification (e.g., *narG*, *nirS*, *nirK*, *nosZ*) had a relatively high abundance detected in all wells except LF2B and MLS35 (Figure S4).

C. SUPPORTING TABLES

Table S1. Other part of geochemistry variables measured in groundwater from each monitoring well

Well	Alk ^a mg/L	Br ⁻ mg/L	Mg ²⁺ mg/L	Fe ²⁺ mg/L	Mn ²⁺ mg/L	Na ⁺ mg/L	Sr ²⁺ mg/L	K ⁺ mg/L	B ³⁺ mg/L	Ba ²⁺ mg/L	Ca mg/L	Si mg/L
LF2B	2948.6	4.6	132.1	11.9	0.6	643.1	5.2	110.8	2.7	13.2	228.3	17.53
MLS35	2494.3	<0.5	118.2	12.5	0.6	635.4	5.2	31.0	2.6	10.0	244.5	16.4
MLS36	2366.3	3.4	ND ^b	ND	ND	ND	ND	ND	ND	ND	ND	ND
MLS37	2636.0	3.9	102.7	14.0	0.6	635.0	4.1	163.3	3.7	7.3	154.9	16.2
MLS38	2425.5	4.4	116.8	11.8	0.4	574.0	5.3	60.8	3.7	9.2	232.7	13.9
MLS54	2133.9	3.8	126.1	16.0	0.5	566.1	5.5	10.5	3.6	4.6	287.9	15.2
MLS55	2008.2	3.7	127.5	10.7	0.6	530.0	5.0	6.6	3.3	1.1	311.5	16.8
MLS80	1697.6	3.6	129.6	8.1	0.7	434.0	4.5	6.0	1.4	1.1	316.1	15.8

^a HCO₃⁻ concentration;

^b No data.

Table S2. Distribution of functional genes detected for major functional gene categories detected by GeoChip 3.0

Gene category	GeoChip 3.0 ^a	LF2B	MLS35	MLS36	MLS37	MLS38	MLS54	MLS55	MLS80
Carbon cycling	5196 (19)	67 (15)	70 (12)	154 (12)	185 (14)	131 (12)	168 (10)	151 (13)	137 (14)
Energy process	508 (2)	8 (2)	10 (2)	26 (2)	25 (2)	23 (2)	37 (2)	25 (2)	25 (2)
Metal Resistance	4870 (18)	109 (24)	144 (25)	297 (24)	303 (24)	245 (23)	406 (25)	260 (23)	241 (24)
Nitrogen	3763 (14)	38 (8)	44 (8)	123 (10)	117 (9)	115 (11)	161 (10)	140 (13)	111 (11)
Organic Remediation	8614 (31)	176 (38)	224 (39)	535 (42)	495 (39)	415 (39)	674 (42)	412 (37)	376 (38)
Phosphorus	599 (2)	18 (4)	16 (3)	29 (2)	26 (2)	25 (2)	33 (2)	31 (3)	18 (2)
Sulphur	1504 (5)	24 (5)	35 (6)	52 (4)	69 (5)	59 (6)	78 (5)	63 (6)	49 (5)
Total	27812	458	574	1262	1280	1060	1621	1120	999

Note: the number in parentheses represents the percentage of the category in total gene numbers.

D. SUPPORTING FIGURES

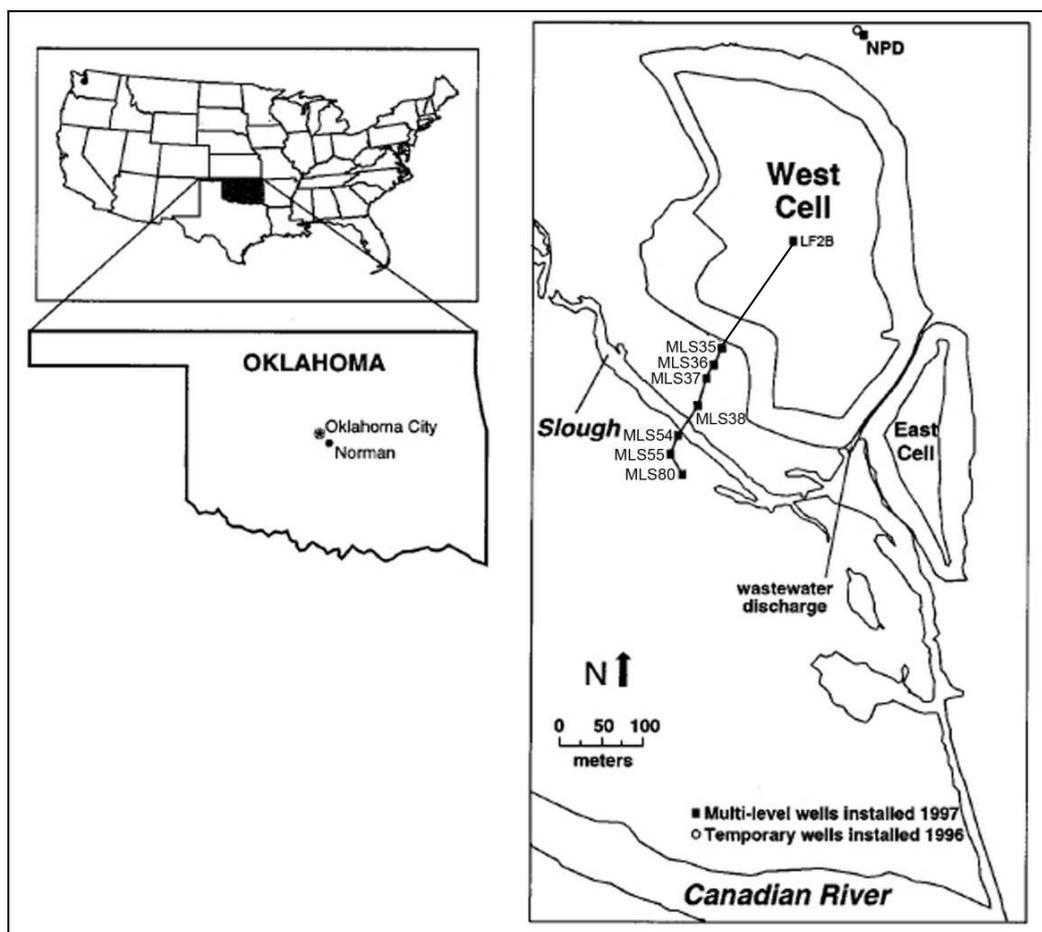
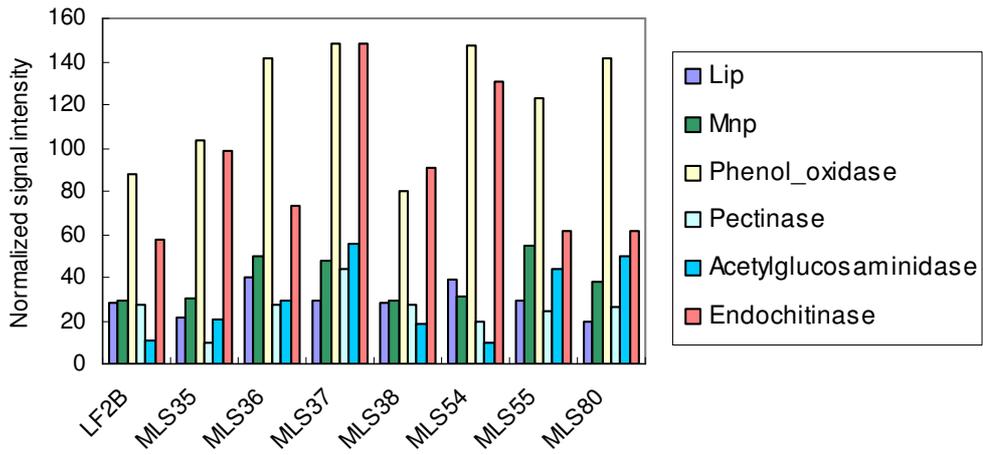
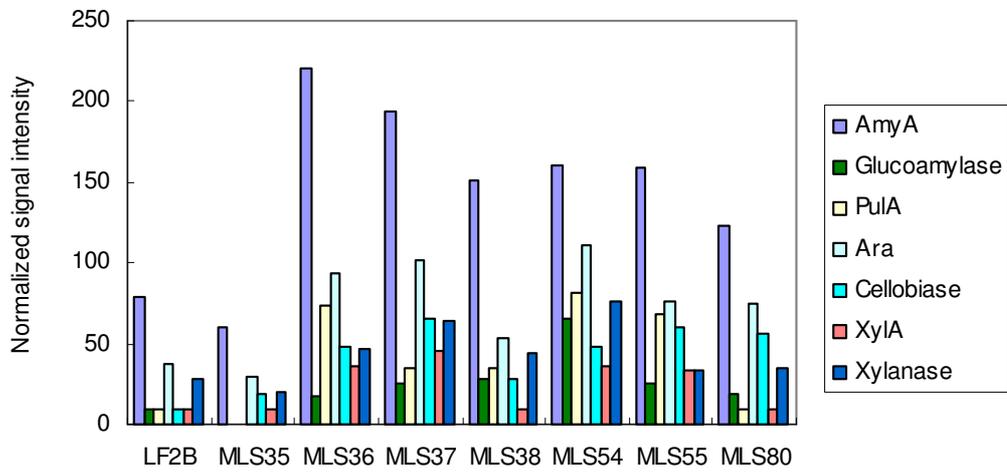


Figure S1. Location of the Norman Landfill and sampling wells. The Norman Landfill is located in Norman, OK, USA. It is unlined (without leachate collection system), and received solid wastes from 1922 until 1985 when it was covered with clay cap and vegetated. Leachate from the unlined landfill has resulted in a groundwater plume that extends downgradient approximately 250m from the landfill toward the Canadian River (approximately 15 m/year) and flows directly beneath the wetland (Scholl and Christenson, 1998; Grossman *et al.*, 2002). Samples were collected from eight monitoring wells (LF2B, MLS35, MLS36, MLS37, MLS38, MLS54, MLS55 and MLS80) shown as diamonds. Numbers refer to wells installed along the groundwater gradient. General site information can be found at <http://csdokokl.cr.usgs.gov/norlan/>. This figure is modified from Cozzarelli *et al.* (2000).



A. Recalcitrant carbon degradation



B. Labile carbon degradation

Figure S2. The normalized signal intensity of genes involved in recalcitrant carbon degradation (A) and labile carbon degradation (B). The signal intensities were the sum of all detected individual gene sequences for each family after log transfer.

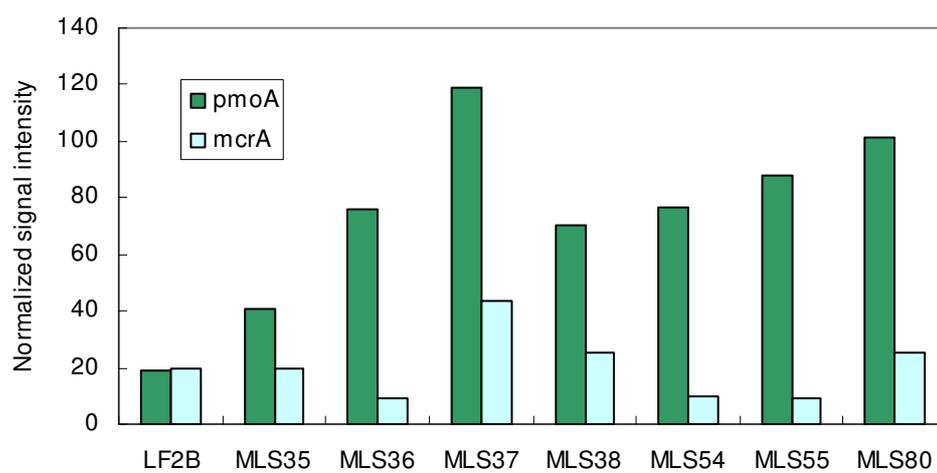


Figure S3. The normalized signal intensity of genes involved in methane production and oxidation. The signal intensities were the sum of all detected individual gene sequences for each family after log transfer.

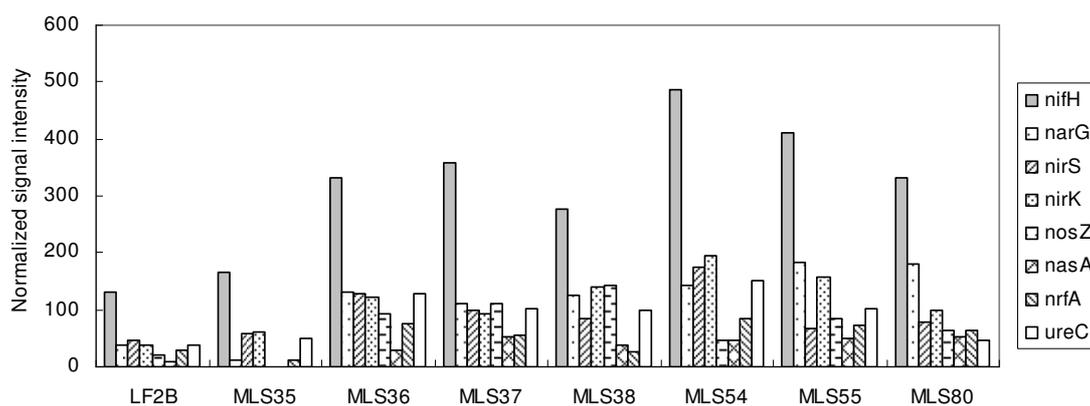


Figure S4. The normalized signal intensity of the major detected gene families involved in the N cycling for the monitoring wells. The signal intensities were the sum of all detected individual gene sequences for each family after log transfer. (1) *nifH*, encoding nitrogenase responsible for N₂ fixation; (2) *narG* encoding nitrate reductase, *nirS* and *nirK*, encoding nitrite reductase; *nosZ*, encoding nitrous oxide reductase, responsible for denitrification; (3) *nasA*, encoding nitrate reductase responsible for assimilatory N reduction; (4) *nrfA*, encoding c-type cytochrome nitrite reductase, responsible for dissimilatory N reduction to ammonium; (5) *ureC*, encoding urease responsible for ammonification.

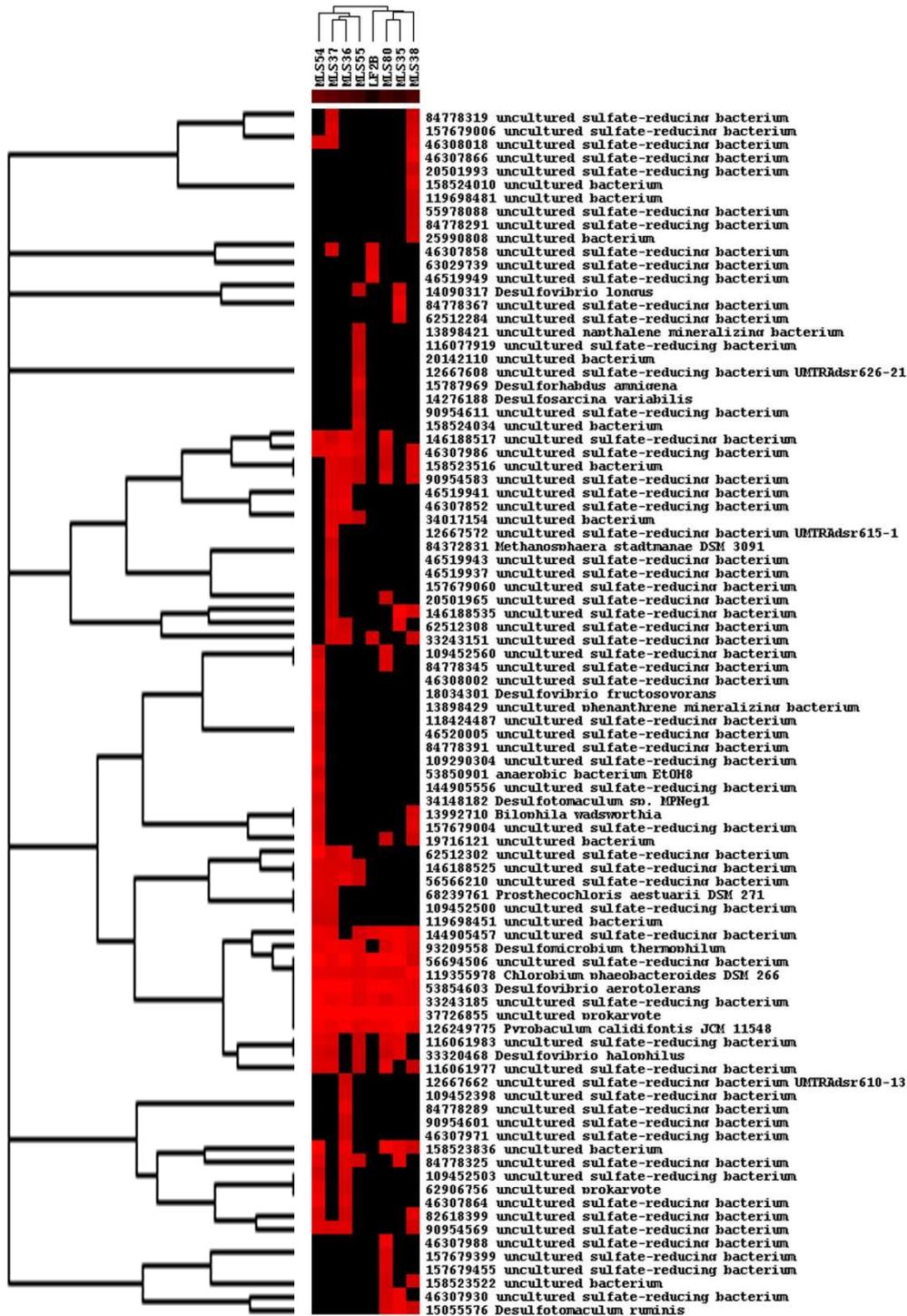


Figure S5. Hierarchical cluster analysis of *dsrA* gene, encoding alpha subunit of dissimilatory sulfite reductase. All genes were used for cluster analysis. Results were generated in CLUSTER and visualized using TREEVIEW. Red indicates signal intensities above background while black indicates signal intensities below background. Brighter red coloring indicates higher signal intensities.

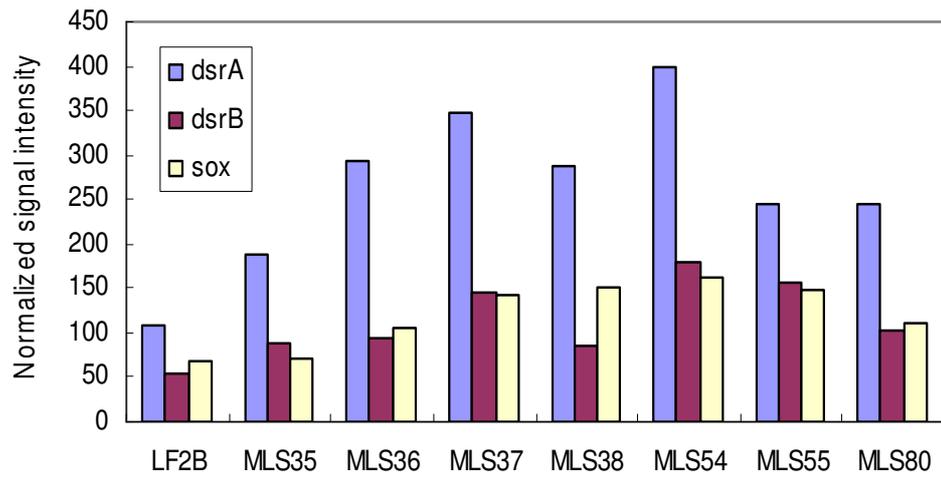


Figure S6. The normalized signal intensity of genes involved in the S cycling for the monitoring wells. The signal intensities were the sum of all detected individual gene sequences for each family after log transfer.

E. SUPPORTING REFERENCES

- [1] Scholl, M. A.; Christenson, S., *Spatial variation in hydraulic conductivity determined by slug tests in the Canadian River alluvium near the Norman Landfill, Norman, Oklahoma*; U.S. Geological Survey Water-Resources Investigations Report 97-4292; U.S. Geological Survey: Reston, Virginia, 1998; <http://pubs.usgs.gov/wri/wri974292/>.
- [2] Wu, L.; Liu, X.; Schadt, C. W.; Zhou, J. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Appl. Environ. Microbiol.* **2006**, *72*, 4931-4941.
- [3] Rhee, S. K.; Liu, X.; Wu, L.; Chong, S. C.; Wan, X.; Zhou, J. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl. Environ. Microbiol.* **2004**, *70*, 4303–4317.
- [4] Tiquia, S. M.; Wu, L.; Chong, S. C.; Passovets, S.; Xu, D.; Xu, Y.; Zhou, J. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **2004**, *36*, 664-670.