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 Zhou2,4,7,8,*

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 µl 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 Templiphi kit (GE Healthcare, Piscataway, NJ, USA) with the following modifications. Spermidine (0.1mM) and single-stranded binding protein (267ng µl⁻¹) were added to improve the amplification efficiency (Wu et al., 2006). 3.0 µg 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 (5xSSC, 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.
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 \textit{nifH} genes encoding a nitrogenase and two \textit{nirK} genes encoding a nitrite reductase responsible for N\textsubscript{2} fixation and denitrification, respectively. A gene encoding a cytochrome c nitrite reductase (\textit{nrfA}) most closely related to the protein responsible for dissimilatory N reduction to ammonium was also detected as were two \textit{ureC} genes encoding a urease responsible for ammonification. Generally, the N-cycling genes, especially N\textsubscript{2} fixation (\textit{nifH}) and denitrification (e.g., \textit{narG}, \textit{nirS}, \textit{nirK}, \textit{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

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

*a* HCO$_3^-$ concentration;

*b* No data.
Table S2. Distribution of functional genes detected for major functional gene categories detected by GeoChip 3.0

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

Note: the number in parentheses represents the percentage of the category in total gene numbers.
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 $\text{N}_2$ 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


