**Functional gene array-based analysis of microbial communities in heavy metals contaminated lake sediments**

**Sanghoon Kang1,\*, Joy D. Van Nostrand1,\*, Heidi L. Gough2, Terry C. Hazen3, David A. Stahl2, Jizhong Zhou1,3,\***

Experimental Procedures

*GeoChip data processing and exploratory analysis*

Three categories of GeoChip data were prepared based on functional relevance to metal contamination: metal resistance and reduction (MRR), sulfate reduction (SR), and carbon cycling (CC). The general response of sediment microbiota to metal contamination was evaluated using several exploratory analyses. Non-metric multidimensional scaling (NMDS) was used for unconstrained ordination. The NMDS results were quantitatively evaluated with analysis of similarity (ANOSIM) ([Clarke and Ainsworth, 1993](#_ENREF_6)). A series of diversity indices were calculated to describe the GeoChip results: Shannon-Weaver diversity index (*H*’), Shannon evenness index (*E*’), Pielou-Kemp transformed Simpson index (-ln(*D*)) and Simpson evenness index (*E*1/*D*). Due to the potential biases in the commonly used Shannon diversity index (*H*’) and the reciprocal of Simpson’s index (1/*D*) ([Rosenzweig, 1995](#_ENREF_37)), the negative log transformation of Simpson’s index (-ln(*D*))was presented as well as the Shannon index ([Pielou, 1975](#_ENREF_35)) for comparison. The Pielou-Kemp transformation (-ln(*D*)) reflects the underlying diversity in a more easily interpretable fashion that is nearly independent of the sample size ([Rosenzweig, 1995](#_ENREF_37)). The response of functional gene diversity to metal contamination was revealed by correlation analysis between pore water metal concentrations and diversity indices.

Functional genes in the three categories were compared separately with concentrations of each pore water metal by multiple correlation analysis. MRR genes were compared with the pore water metal species associated with the resistance gene, then MRR, SR and CCG genes were compared with all available pore water metals. The statistical significance was adjusted by both Bonferroni correction (*α*’ = *α*/*k*) and positive false discovery rate (*p*FDR) ([Storey, 2003](#_ENREF_44)) to account for family-wise error rate (FWER).

**Gradient analysis and multiple regression analysis**.

The spatial patterns of extant environmental microbial communities are usually related to current and historic environmental changes ([Martiny et al., 2006](#_ENREF_31); [O'Malley, 2008](#_ENREF_32)). We used canonical correspondence analysis (CCA), Mantel test and variation partitioning analysis (VPA) to evaluate two-factor relationships. Even though ecological data is often assumed to be unimodal, we evaluated the data structure by comparing the results from redundancy analysis (RDA) and CCA between GeoChip datasets and pore water metals. The results clearly proved a unimodal data structure, thus CCA was used to link microbial communities and pore water metals.

The Mantel test and multivariate correlogram analysis were performed to assess the contribution of geographic distances between sampling sites, which may also have a strong historical component ([Martiny et al., 2006](#_ENREF_31)). The results of CCA and partial CCA were also used to partition the variations of microbial community structure between pore water metals and geographic distances ([Borcard et al., 1992](#_ENREF_5)).

The genetic structure of individual samples were analyzed using multiple regression on each functional gene with the pore water metals using the base function “lm(stats)” of R (v.2.6.1). The independent variables were first tested by collinearity diagnostics (VIF) to improve the stability of the regression model. Models were selected by a stepwise method based on AIC (Akaike Information Criterion) using the function “step(MASS)”. Test of parametric statistics assumptions were performed to verify results along with the linearity assumption for linear regression. Finally the results were summarized by significant *P* values with *p*FDR.

GeoChip interrogated genes were summarized into 10 functional gene categories (C degradation, C fixation, sulfate reduction, metal reduction, methane oxidation, methane generation, N fixation, nitrification, N reduction and organic remediation) to identify functional gene categories significantly influenced by pore water metals. Functional genes belonging to MRRGs (As, Cd, Cr, Cu and Zn) and CCGs (cellulase, chitinase, laccase, mannanase, *pgl*, polygalacturonase, *acs*A, *coo*F, FTHFS, *rbc*L and *rbc*S) were further summarized at the functional gene subgroup level. Multiple regression, Mantel test and CCA were used the same way as described above in identifying important pore water metals.

**Supplemental Results**

**Analysis of three key functional gene categories**

*(i) Metal resistance and reduction*. There were several genes that were significantly correlated with corresponding pore water metal concentrations based on Bonferroni correction and positive false discovery rate (*p*FDR) (Table S2). Four Cu resistance genes were significantly correlated with the pore water Cu concentration. Two Zn resistance genes, one As resistant gene and one Cd resistance gene were significantly correlated with corresponding pore water metal concentrations. Among the 13 correlated genes, only a Cr resistance gene (31794232, *Mycobacterium bovis* AF2122/97) was negatively correlated with the Cr concentration in pore water (*r* = -0.207, *P* = 0.037).

Since functional genes often confer resistant to multiple heavy metals, each pore water metal concentration was compared with all detected MRR genes (Table S3). The list of significant MRR genes with As, Cd, Cr, Pb or Zn pore water concentrations were quite similar. All of these metals shared a majority of the genes with significant correlations. For example, the lists for As and Zn were almost identical with quite similar correlation coefficients, which indicates a very significant correlation between these two metals (*r* = 0.990, *P* < 0.001). The list of MRR genes with significant correlations with Cu and Mn were different from the rest. Cu concentration was correlated with 34 MRR genes, and 2 and 11 of them were significant after Bonferroni correction and *p*FDR (*q* < 0.1), respectively. Only two of them shared with other metals, which were resistant to multiple metals (17548710, 26989131-cobalt, Zn and Cd). Mn was correlated with 15 MRR genes, which is the second most, but none of them was significant after family-wise error rate (FWER) correction. A mercury resistant gene (1143575, *Shewanella putrefaciens*) was the only gene shared with other metal (Pb). Eleven MRR genes were negatively correlated with Mn, while only one other gene was negatively correlated (7328307-cobalt, nickel) with other metals (Zn and Pb). All 11 selected genes correlated with PbPb became insignificant after FWER correction. Several MRR genes were consistently correlated with most pore water metal concentrations. For instance 17548710 (*Ralstonia solanacearum* GMI1000) and 26989131 (*Pseudomonas putida* KT2440) were significantly correlated with all but Mn. There were nine other MRR genes correlated with all but Cu and Mn as well (168919, 15610714, 1785478, 17548710, 2314827, 26989131, 33592181, 33600180 and 37520133).

Five genes were significant both by the correlation with pore water metal concentrations and multiple regression with pore water Cu concentration: 151189-*Pseudomonas syringae*,16763879-*Salmonella typhimurium* LT2,17548875-*Ralstonia solanacearum* GM1000, 17937687-*Agrobacterium tumefaciens* str. C58 and 21244355-*Xanthomonas axonopodis* pv. citri st. 306. MRR genes were detected less than other gene categories even if the samples were from heavy metal contaminated sediment and it was one of the largest functional gene categories. Cu reduction genes were most significantly correlated with pore water Cu concentrations, among other functional genes compared with corresponding metals.

*(ii) Sulfate reduction.* SR genes were further analyzed since many sulfate reducing bacteria (SRB) can also reduce various metals, and a previous study from Lake DePue indicated a negative correlation between the sulfate reducing rate and pore water metal concentrations {Gough, 2008 #55}. SR genes is one of the smaller functional gene categories in GeoChip 2.0 (701 genes, 6.68% of all genes targeted), but a higher percentage of SR genes were detected (20.3%). The lists of SR genes of significant correlations with As, Cd, Cr, Pb or Zn pore water concentrations were quite similar as was in MRR genes. The list of SR genes with significant correlations with Cu and Mn were different from the rest (Table S4). Cu was correlated with 18 SR genes, and 3 and 10 of them were significant after Bonferroni correction and *p*FDR (*q* < 0.1), respectively.Only 2 of the 18 SR genes were shared with other metals. Mn, on the other hand, was correlated with just six SR genes, which were uniquely correlated with Mn only, although all of them were not significant after FWER correction. In addition, four of them were negatively correlated with Mn, while there was only one other gene negatively correlated (15193497, *dsr*A) with other metals (As, Zn and Pb). Several genes were consistently correlated with different metals; FW300167B (*dsr*B, lab clone) was significant with all but Mn. Five genes were significant with all but Cu and Mn: 21673681, FW015014A, 19716123, 25990790 and 14389163. Most SR genes were from either environmental clones or uncultured bacteria. There were only five genes from pure culture in the genus of *Desulfotomaculum* and *Desulfovibrio*. A *dsr*A gene (21673681) from thermophilic green-sulfur bacterium (*Chlorobium tepidum* TLS) was one of the genes significantly correlated with most pore water metals.

Unlike MRR genes, most genes significantly regressed were from lab clones. There was only one gene both significantly correlated and regressed with pore water Zn concentration (14389129-*dsr*B-uncultured sulfate-reducing bacterium). As a small category SR genes were detected above the average indicating an active involvement of SRB in metal reduction, however very few SR genes were from pure culture SRB.

*(iii) Carbon cycling*. CC genes were chosen for further analysis because of the observation of a rough correlation between organic carbon and metal concentrations, and a negative correlation between microbial biomass and total organic carbon (TOC) {Gough, 2008 #56}. The detection rate of carbon degradation and fixation genes was little higher (14.9%) than MRR genes but less than all functional genes together. Like MRR genes and SR genes, the lists of genes with significant correlations with As, Cd, Cr, Pb or Zn were similar (Table S5). Cu had the most correlated CC genes (27), while Mn had only six. 8 and 19 of those 27 CC genes were significant by Bonferroni correction and *p*FDR respectively, but none of the 6 genes that were correlated with Mn were significant by FWER correction. Also like MRR genes and SR genes, CC genes correlated with Cu and Mn were mostly unique to Cu and Mn. Cu shared four genes (662361, 15823752, 16764956-cellulcase and 19697890-laccase) with other metals, while Mn shared only one gene with Pb (8926979-cellulase). A cellulase gene (4732055) from *Rhizobium leguminosarum* and an *rbc*S gene (1850940) from *Synechococcus* species were the two genes correlated with only Cu and Mn. Four of the six CC genes selected with Mn were negatively correlated just like in SR genes. An *rbc*S gene (17368193, *Sinorhizobium meliloti*) wasthe only CCG negatively correlated with other than Mn (As, Pb and Zn). Unlike SR genes, a majority of selected genes were from isolates, but seven of 16 carbon fixation genes (mostly *rbc*L gene) were from uncultured bacteria.

Four CC genes were both significantly correlated and regressed with pore water Cr concentration: 80483-cellulase-*Clostridium acetobutylicum*, 121841-cellulase-*C. saccharobutylicum*, 15419704-cellulase-*Ralstonia solanacearum* and 25900624-FTHFS-*Desulfovibrio baarsii*. With pore water Cu, eightCC genes were identified: cellulase-15823752-*Bacillus circulans*, 16764956-*Salmonella typhimurium* LT2, 15054476-*Aspergillus kawachii*; chitinase-19526733-uncultured bacterium, 19526729-uncultured bacterium; laccase-37359391-*Rigidoporus microporus*, 578092-*Phlebia radiata*; polygalacturonase-34366094-*A. aculeatus*. Majority of CC genes correlated with pore water metal concentration were carbon degradation genes mostly cellulose, chitinase and laccase.