

1 **Supplemental Material for:** Applied and Environmental Microbiology

2

3 **New arsenate reductase gene (*arrA*) PCR primers for diversity assessment**  
4 **and quantification in environmental samples**

5

6 Babur S. Mirza<sup>1\*</sup>, Darwin L. Sorensen<sup>1</sup>, R. Ryan Dupont<sup>1,2</sup>, Joan E. McLean<sup>1,2\*</sup>

7

8

9

10 <sup>1</sup>Utah Water Research Laboratory, Utah State University, Logan, UT 84322-8200,  
11 USA

12 <sup>2</sup>Department of Civil and Environmental Engineering, Utah State University, Logan,  
13 UT 84322-8200, USA

14

15 **Keywords:** Arsenate reductase gene (*arrA*), High throughput sequencing, qPCR of  
16 *arrA* gene

17

18

19

20 **Address correspondence to:** Babur S. Mirza, babur.mirza@usu.edu, or

21

Joan E. Mclean, joan.mclean@usu.edu

22

<sup>1</sup>Utah State University

23

Utah Water Research Laboratory

24

8200 Old Main Hill,

25

Logan UT 84322-8200.

26

Tel: (435) 797-3199

27

Fax: (435) 797 3663

28

29

30

31

32 **Supplementary Figures:**

33 **Fig. S1:** Maximum-likelihood phylogenetic tree based on partial sequences (630 bp)  
34 of the arsenate reductase genes (*arrA*) of the 54 *arrA* gene clones previously  
35 generated (Mirza et al., 2014) and of 463 *arrA* gene sequences both from cultured  
36 and uncultured organisms from GenBank. These 54 *arrA* gene clones were  
37 evaluated for the successful amplification of this gene using the newly developed  
38 PCR primers. Sequences from other studies were collapsed into OTUs (at 90%  
39 identity) to show the overall relationships of these clusters to our clone sequences.  
40 Numbers at branch nodes reflect the maximum likelihood bootstrap support values.  
41 The *psrA* gene from *Wolinella succinogenes* (X65042) was used as an out-group.

42 **Fig. S2:** (A) The DNA sequence identities of the newly developed *arrA* gene primers  
43 with the *arrA* gene sequences from five different pure culture isolates. (B) Overall  
44 *arrA* gene region showing the presence of the conserved region with these five  
45 isolates at the newly developed primer sites.

46 **Fig. S3:** Translated amino acid sequence identities of the newly developed *arrA* gene  
47 primers with the *arrA* gene sequences from different pure culture isolates.

48 **Fig. S4:** PCR amplification of *ArrA* gene using the newly developed PCR primers  
49 from *arrA* gene clones, soil and groundwater samples. DNA ladder (1kb) was  
50 presented in the right most lane.

51 **Fig. S5:** Maximum-likelihood phylogenetic analysis of translated amino acids of  
52 partial *arrA* gene sequences along with the other arsenite oxidase and Mo-containing  
53 proteins sequences.

54 **Fig. S6:** Multidimensional scaling plot based on the Bray-Curtis similarity index at  
55 90% *arrA* gene identity showed differences in the arsenate reducing microbial  
56 species between soil and groundwater well samples from Cache Valley sites.

57 **Fig. S7:** Nonmetric multidimensional scaling plots representing different groundwater  
58 quality parameters of the Cache Valley Basin (CVB) domestic wells based on the  
59 Euclidean distance matrix. The four samples that have been associated with the

60 Logan landfill site were not included from this analysis. Length and direction of  
61 arrows indicate the strength and nature of association of different water quality  
62 parameters, respectively.

63

#### 64 **Supplemental Table:**

65 **Table S1:** Sampling locations, arsenic concentrations, and other chemical  
66 characteristics of soil and groundwater samples used in this study.

67 **Table S2:** *ArrA* gene forward and reverse primers (black) along with sequencing  
68 adapter (blue) and unique identifier sequences (green).

69 **Table S3:** Distribution of the previously described *arrA* gene sequences from the  
70 GenBank among different major phylogenetic groups.

71 **Table S4:** Distribution of the previously described *arrA* gene sequences (9) from the  
72 Cache Valley Basin, Utah.

73

#### 74 **REFERENCES**

- 75 **1. Ying SC, Damashek J, Fendorf S, Francis CA.** 2015. Indigenous arsenic(V)-  
76 reducing microbial communities in redox-fluctuating near-surface sediments of  
77 the Mekong Delta. *Geobiol* **13**:581–587.
- 78 **2. Zhang SY, Zhao FJ, Sun GX, Su JQ, Yang XR, Li H, Zhu YG.** 2015. Diversity  
79 and Abundance of Arsenic Biotransformation Genes in Paddy Soils from  
80 Southern China. *Environ Sci Technol* **49**:4138-4146.
- 81 **3. Ohtsuka T, Yamaguchi N, Makino T, Sakurai K, Kimura K, Kudo K, Homma**  
82 **E, Dong DT, Amachi S.** 2013. Arsenic dissolution from Japanese paddy soil by  
83 a dissimilatory arsenate-reducing bacterium *Geobacter* sp. OR-1. *Environ Sci*  
84 *Technol* **47**:6263-6271.
- 85 **4. Mumford AC, Barringer JL, Benzel WM, Reilly PA, Young LY.** 2012.  
86 Microbial transformations of arsenic: mobilization from glauconitic sediments to  
87 water. *Water Res* **46**:2859–2868.
- 88 **5. Song B, Chyun E, Jaffé PR, Ward BB.** 2009. Molecular methods to detect and  
89 monitor dissimilatory arsenate-respiring bacteria (DARB) in sediments. *FEMS*  
90 *Microbiol Ecol* **68**:108–117.
- 91 **6. Upadhyaya G, Clancy TM, Brown J, Hayes KF, Raskin L.** 2012. Optimization  
92 of arsenic removal water treatment system through characterization of terminal  
93 electron accepting processes. *Environ Sci Technol* **46**:11702-11709.

- 94       **7. Héry M, Rizoulis A, Sanguin H, Cooke DA, Pancost RD, Polya DA, Lloyd**  
95       **JR.** 2015. Microbial ecology of arsenic-mobilizing Cambodian sediments:  
96       lithological controls uncovered by stable-isotope probing. *Enviro Microbiol*  
97       **17**:1857–1869.
- 98       **8. Héry M, YanDongen BE, Gill F, Mondal D, Vaughan DJ, Pancost RD, Polya**  
99       **DA, Lloyd JR.** 2010. Arsenic release and attenuation in low organic carbon  
100       aquifer sediments from West Bengal. *Geobiol* **8**:155-168.
- 101       **9. Mirza BS, Muruganandam S, Meng X, Sorensen DL, Dupont RR,**  
102       **McLean JE.** 2014. Arsenic(V) reduction in relation to Iron(III)  
103       transformation and molecular characterization of the structural and  
104       functional microbial community in sediments of a basin-fill aquifer in  
105       Northern Utah. *Appl Environ Microbiol* **80**:3198–3208.
- 106
- 107

Fig. S1

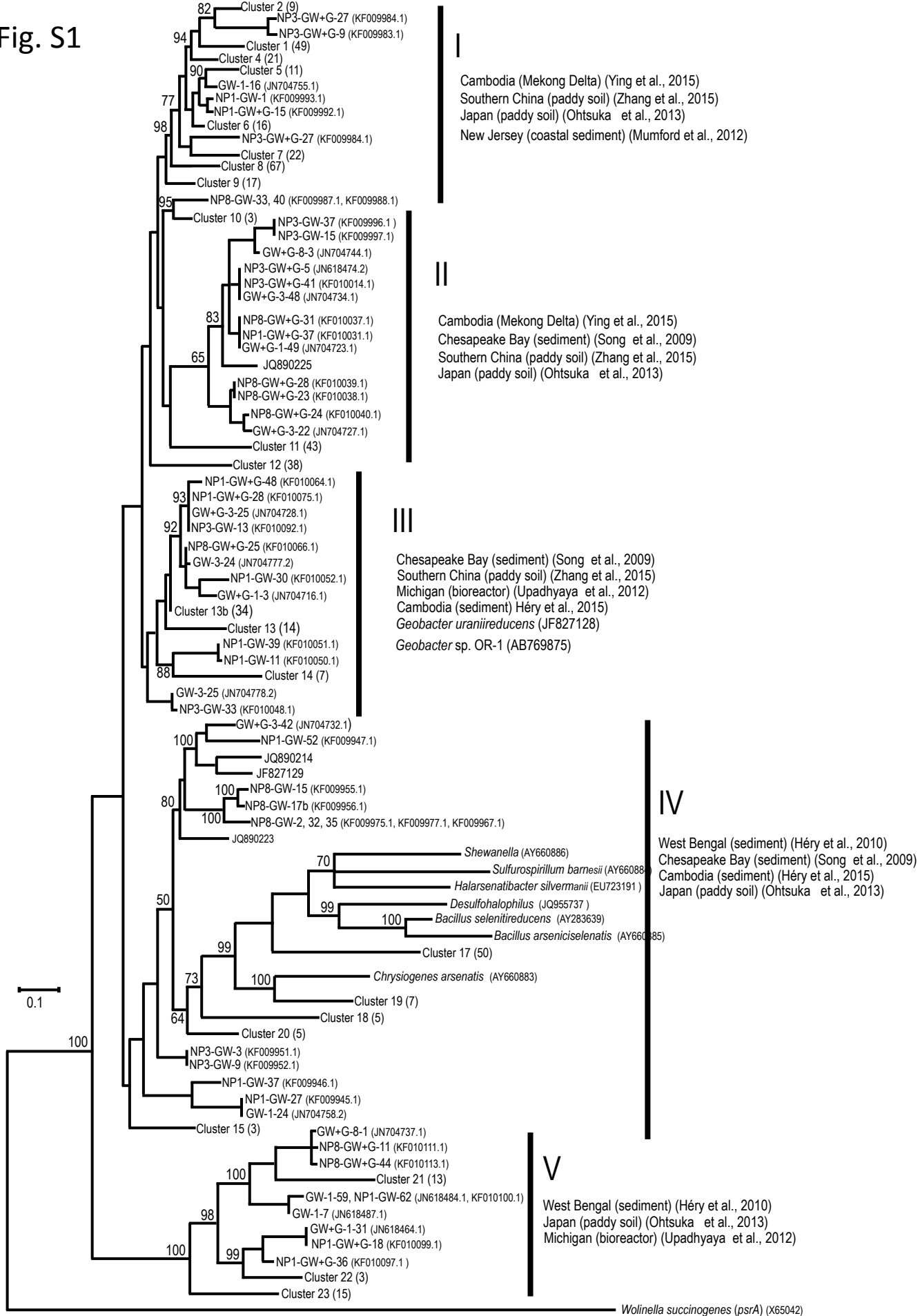


Fig. S2A

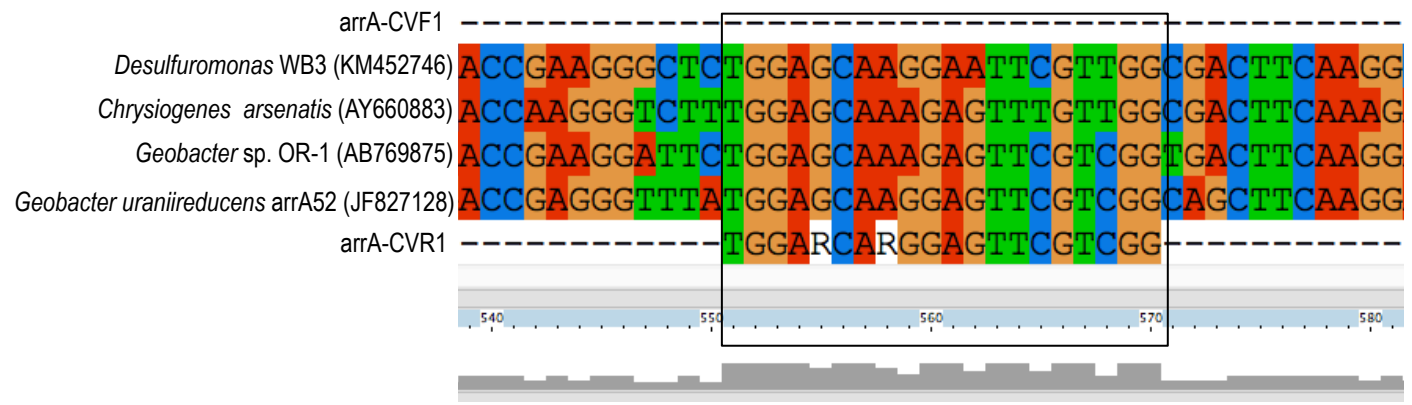
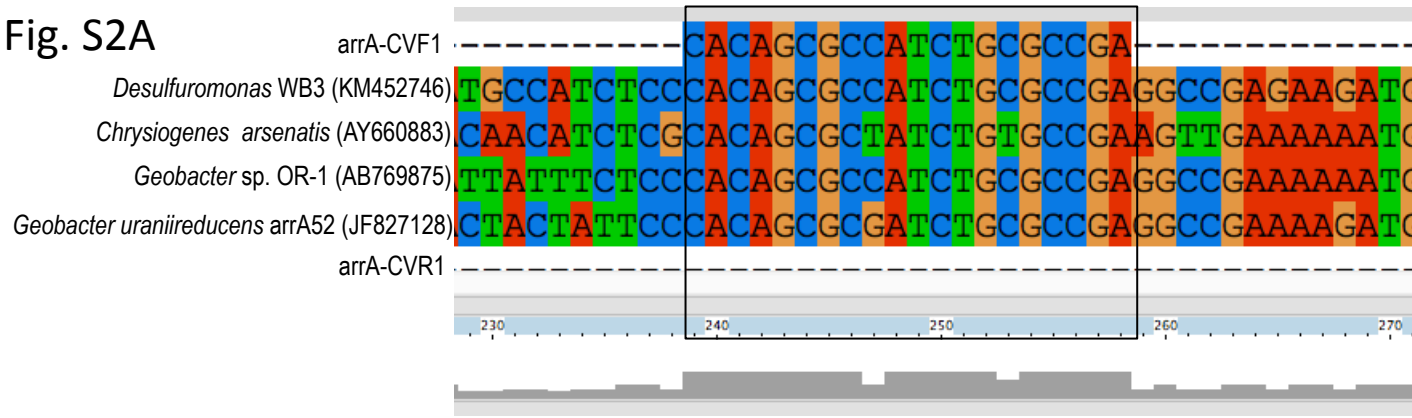


Fig. S2B

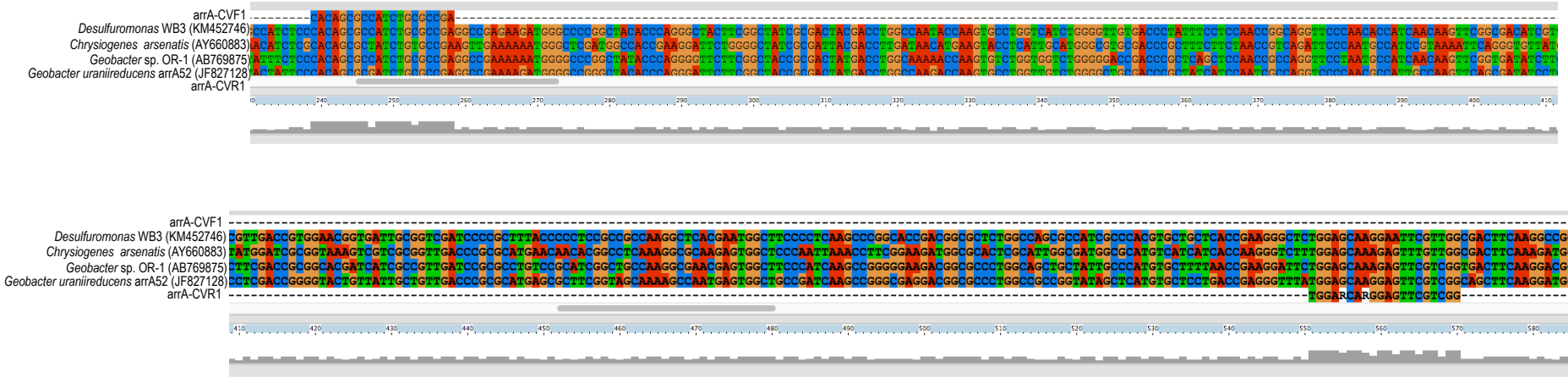


Fig. S3

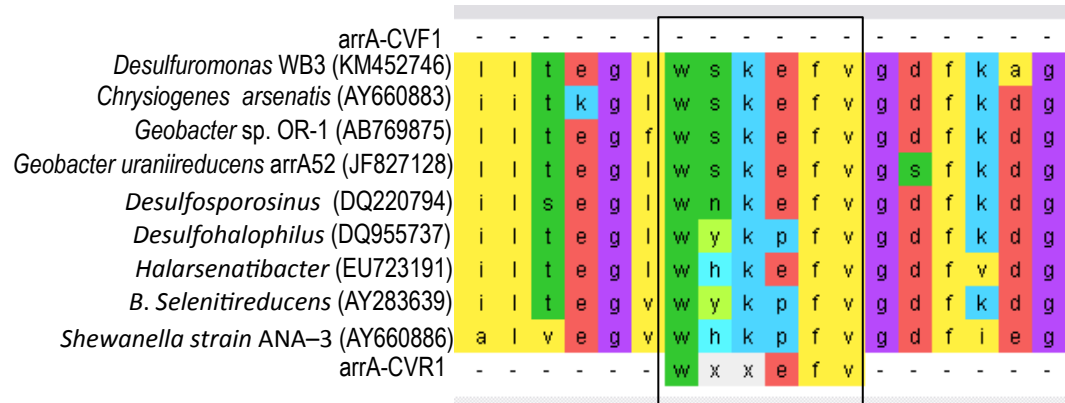
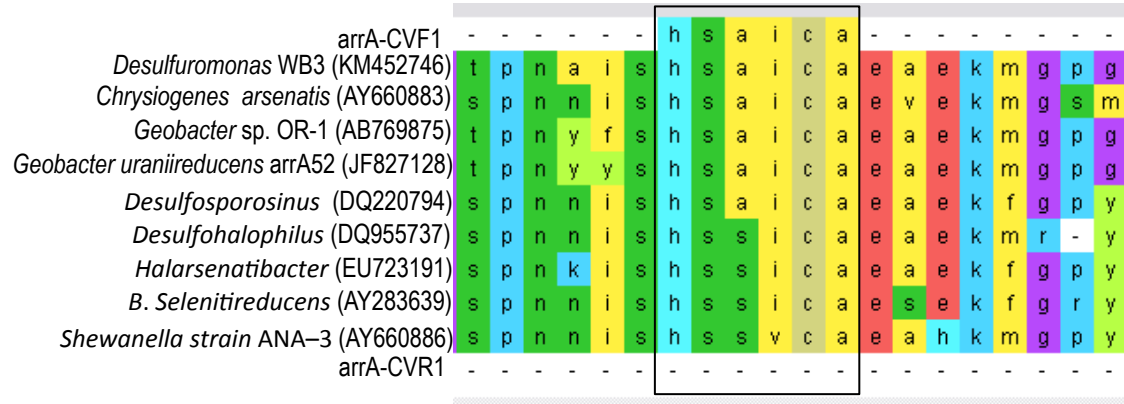


Fig. S4

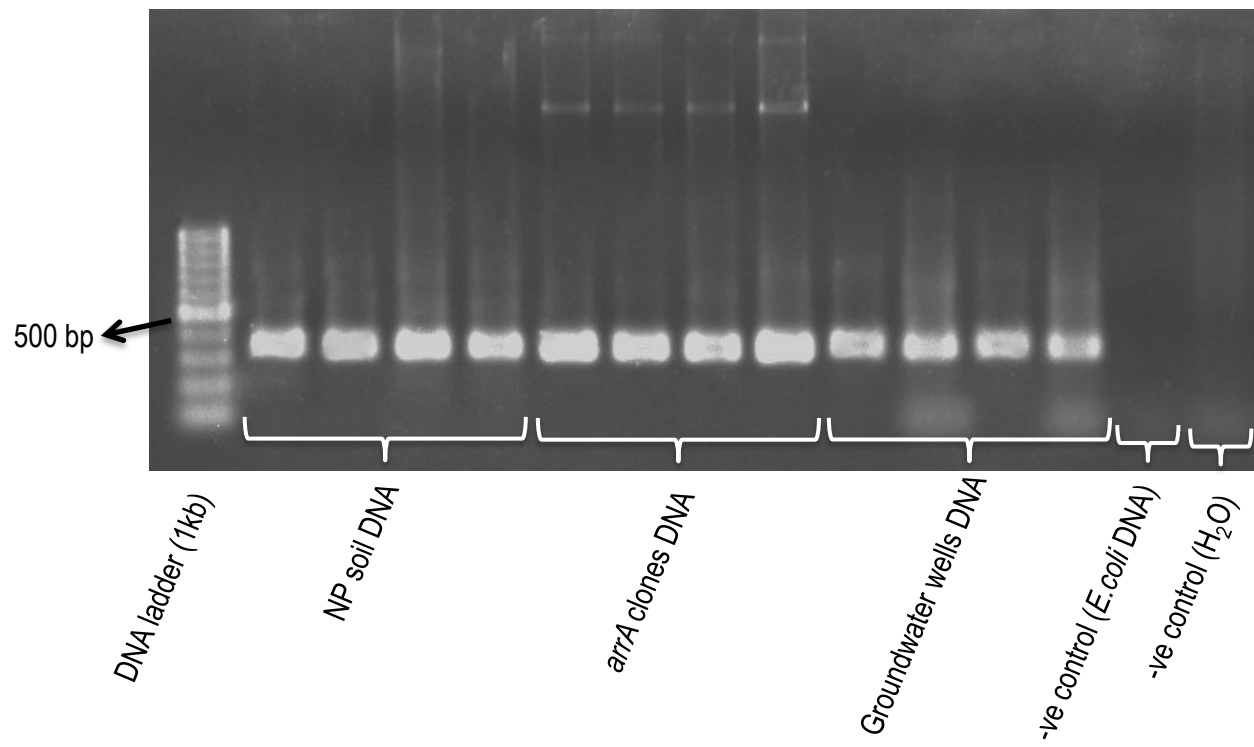
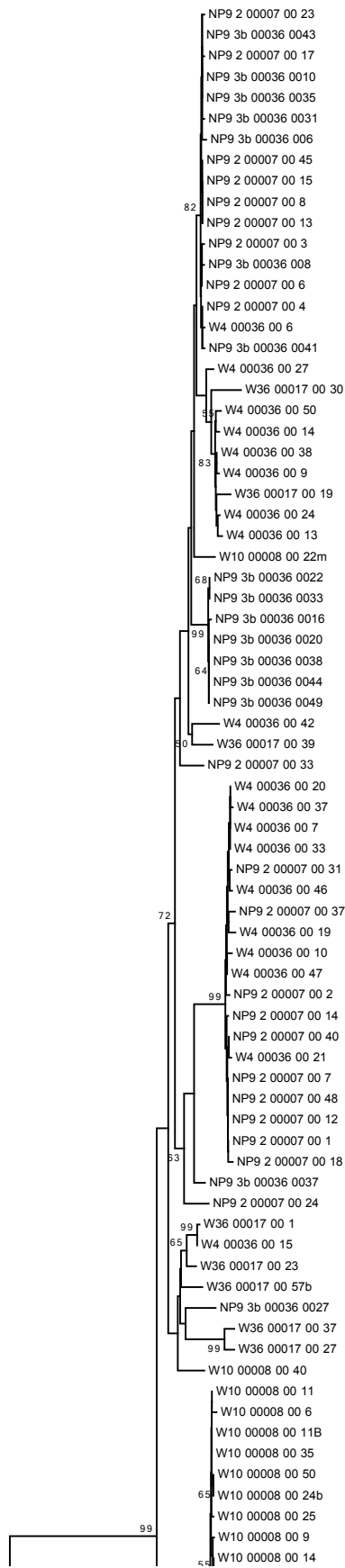




Fig. S5



Continue Fig. S5

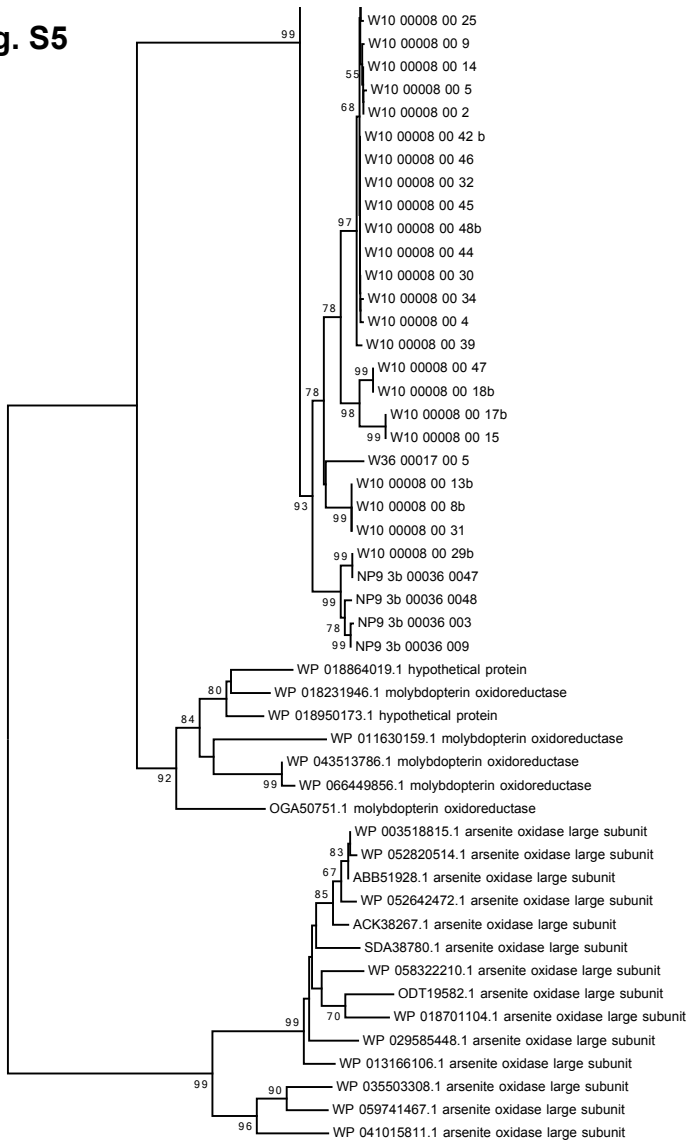


Fig. S6

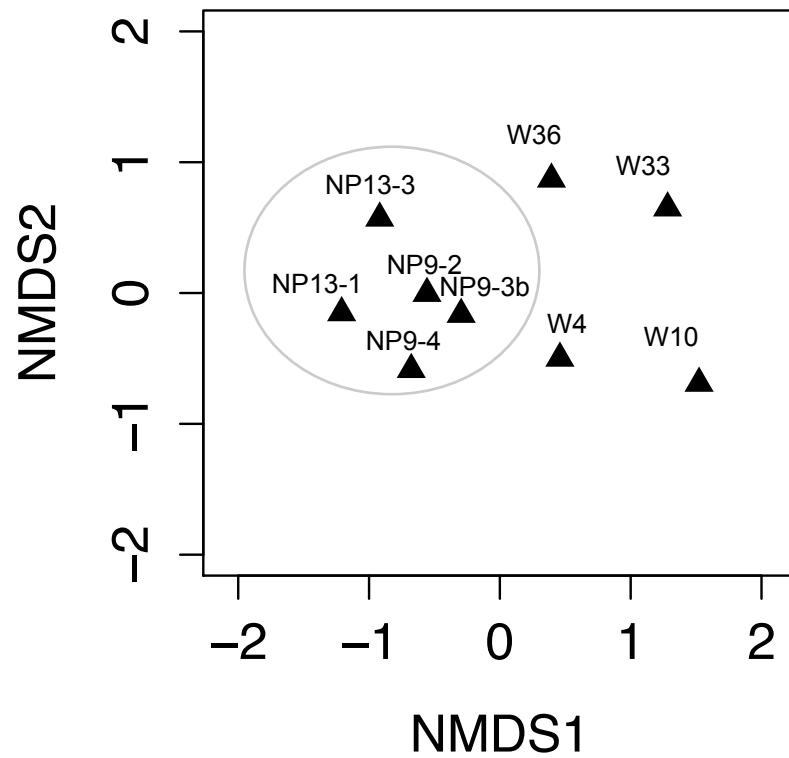
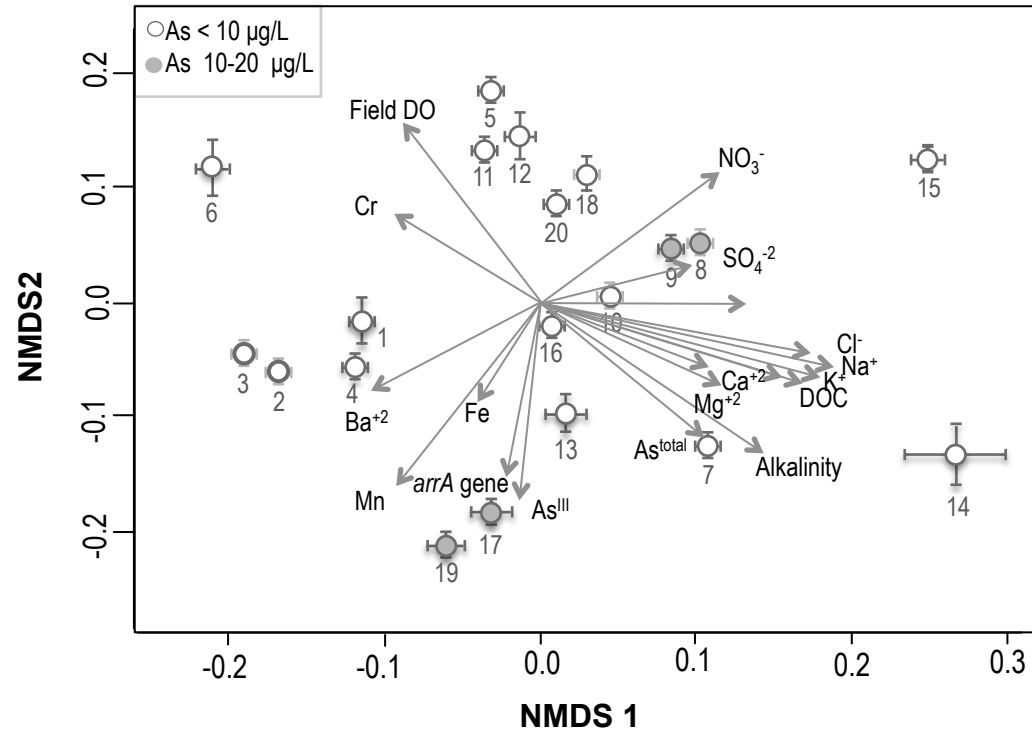


Fig. S7



**Table S1: Sampling locations, arsenic concentrations, and other chemical characteristics of soil and groundwater samples used in this study (reported as mean values).**

Wells number	Locations	Depth cm	As <sup>III</sup> in µg/L (µM)	*As <sup>total</sup> in µg/L (µM)	arrA gene (log)	DO mg/L	EC µS/cm	pH	DOC mg/L	Alkalinity mg CaCO <sub>3</sub> /L	Cl mg/L	SO <sub>4</sub> <sup>2-</sup> (mg/L)	NO <sub>3</sub> -N (mg/L)	Na (mg/L)	Ca (mg/L)	Mg (mg/L)	K (mg/L)	Fe (mg/L)	Mn (µg/L)	Ba (µg/L)	Cr (µg/L)
1	41°40.081'N 111°54.602'W	220	0.7 (0.009)	1.2 (0.016)	3.5	6.5	509	7.01	0.9	309	13.6	24.9	0.53	18.0	68.6	32.4	6.7	0.66	478	358	2.2
2	41°40.042'N 111°54.322'W	168	0.5 (0.007)	0.7 (0.01)	3.3	5.3	557	7.39	0.4	251	10.4	62.6	0.01	9.6	72.8	30.2	3.7	3.667	567	240	2.1
3	41°40.183'N 111°54.483'W	160	0.4 (0.005)	0.5 (0.01)	3.4	2.8	573	7.31	0.3	245	10.1	58.7	0.01	8.4	72.8	28.8	2.5	4.30	354	190	3.7
4	41°40.182'N 111°54.840'W	200	1.9 (0.025)	2.5 (0.03)	4.7	6.2	610	7.16	0.4	323	17.1	3.8	0.02	19.5	69.8	31.2	5.9	1.95	357	225	1.9
5	41°41.814'N 111°58.660'W	237	0.1 (0.001)	1.0 (0.01)	3.5	7.5	420	7.10	0.4	222	20.2	6.4	2.06	20.4	74.1	8.6	1.9	0.04	2.6	84	2.1
6	41°41.572'N 111°48.969'W	450	0.1 (0.001)	1.9 (0.03)	4.2	8.2	417	7.18	1.1	178	19.3	2.5	0.57	9.5	60.9	7.7	1.1	9.95	112	58	10.8
7	41°44.939'N 111°57.804'W	245	5.4 (0.072)	6.4 (0.09)	5.4	5.7	916	7.31	2.7	430	41.8	0.9	0.02	137	33.2	20.0	7.2	0.79	87	139	1.7
8	41°46.108'N 111°59.040'W	92	0.1 (0.001)	12 (0.16)	4.0	6.2	815	7.11	1.4	310	64.4	66.0	2.39	57.7	83.3	34.2	8.9	0.56	4.6	87	2.3
9	41°46.141'N 111°59.305'W	150	0.1 (0.001)	10.9 (0.15)	3.8	5.6	804	7.18	1.1	288	57.3	69.0	2.14	66.1	74.6	28.3	8.9	0.12	2.6	44	2.2
10	41.76946 N -112.00803 W	298	0.7 (0.009)	9.4 (0.13)	4.8	3.9	694	7.23	1.2	247	48.7	50.2	0.25	34.8	75.6	27.0	5.8	0.34	28.8	118	1.9
11	41°48.406'N 111°52.020'W	292	0.1 (0.001)	4.1 (0.05)	4.3	6.4	727	7.25	0.3	191	15.5	5.6	0.18	24.8	34.8	18.2	2.0	0.028	0.8	115	2.4
12	41°49.065'N 111°51.177'W	200	0.1 (0.001)	1.1 (0.01)	3.1	7.9	570	7.07	0.8	236	9.7	9.8	4.08	18.9	52.3	24.7	4.5	0.020	0.3	104	3.2
13	41°47.716'N 111°52.890'W	150	3.3 (0.044)	4.2 (0.06)	4.8	4.3	457	7.51	1.4	220	8.5	0.3	0.06	31.7	40.5	14.8	1.5	2.42	51.7	75	1.9
14	41°56.021'N 112°02.710'W	100	1.0 (0.013)	8.2 (0.11)	3.4	2.5	1585	7.27	3.3	519	187	29.3	0.02	172	62.0	46.3	20.0	9.66	206	116	3.1
15	41°57.082'N 111°58.441'W		0.2 (0.003)	7.9 (0.11)	3.2	5.3	686	7.12	2.3	294	51.9	146	13.76	82.7	82.4	40.1	11.5	0.30	0.6	37	2.1
16	41°54.959' N 112° 3.605' W	183	1.6 (0.021)	8.7 (0.12)	5.0	4.2	586	7.12	0.8	306	15.6	51.4	0.22	15.9	90.3	22.9	3.3	1.02	57.7	32	1.6
17	41°55.042'N 112°3.626'W	150	12.6 (0.168)	15.5 (0.21)	5.3	7.0	618	7.02	0.6	281	13.1	16.3	0.18	9.6	77.3	20.3	5.8	5.13	171	42	2.0
18	41°54.267'N 112°2.718'W	140	0.10 (0.001)	6.9 (0.09)	4.3	3.1	645	7.20	0.6	249	21.8	23.5	1.87	15.5	93.4	11.9	5.3	0.024	0.4	85	2.2
19	41°54.187'N 112°3.062'W	57	15.6 (0.208)	18.2 (0.24)	5.7	3.1	645	7.20	0.9	275	22.7	29.0	0.09	21.0	86.4	16.5	6.4	1.05	1109	158	1.9
20	41°54.325'N 112°2.595'W	130	0.2 (0.003)	6.6 (0.09)	3.0	0.6	4120	7.20	0.8	265	23.4	20.7	0.49	14.5	93.1	13.1	5.5	0.045	2.6	87	2.9
33	41°44.044'N 111°52.778' W	6.8	50.5 (0.674)	87.1 (1.16)	8.2	0.6	3840	6.81	44.9	1265	706	4.9	0.35	354	178	181	39.2	5.30	1340	914	1.6
34	41°43.838'N -111°52.792'W	8	16.9 (0.226)	23.6 (0.31)	6.6	0.2	8230	6.95	28.2	895	631	21.7	0.02	440	105	147	46.3	1.24	1179	457	0.7
35	41°43.755'N -111° 52.790'W	6.6	30.0 (0.40)	20.7 (0.28)	6.6	0.6	3330	7.40	4.0	451	130	5.8	0.78	87.3	54.4	59.4	11.8	0.069	8	170	0.3
36	41°44.028'N 111°52.445' W	4.2	1.3 (0.017)	85.2 (1.14)	6.0	6.5	509	7.84	10	1005	424	155	0.10	591	19.4	67.1	27.0	0.039	40	160	
<b>Soil pore water samples</b>																					
NP9-2	41°44.074'N					ND									17.0						
	111°52.3707'W	38	4.71	10.5 (0.14)	ND**		315	7.71	13.7	119	108	92.2	1.9	124		10.9	4.26	0.012	2.77	325	ND
NP9-3	41°44.074'N					ND									7.7						
	111°52.3707'W	63	5.60	23.2 (0.31)	ND		331	7.89	17.3	151	96.0	74.1	0.08	135		7.0	5.56	0.04	1.87	196	ND
NP9-4	41°44.074'N					ND									5.04						
	111°52.3707'W	93	45.4	158. (2.11)	ND		290	8.02	6.41	198	57.4	72.4	0.14	134		4.81	6.90	0.02	2.1	145	ND
NP13-1	41°43.992'N 111°52.032'W	11	1.21	17.2 (0.23)	ND	ND	783	7.96	15.8	394	146	250	0.06	323	10.1	4.10	4.11	0.58	32.9	85.0	ND
NP13-3	41°43.992'N 111°52.032'W	62	32.0	118. (1.57)	ND	ND	661	8.46	11.4	230	170	106	0.17	214	5.35	18.3	11.3	0.07	1.64	268	ND

DO (dissolved oxygen), EC (electrical conductivity), DOC (dissolved organic carbon)

\*For soils: concentration is in a water extraction of the soil (soil pore water)

\*\*not determined

**Table S2. *ArrA* gene forward and reverse primers (black) along with sequencing adapter (blue) and unique identifier sequences (green).**

Primer number	<i>ArrA</i> gene primer sequences along with the 454 sequencing adaptor and unique identifiers
arrA-CVF1 (1)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACATGACGAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (2)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACGACAGCTC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (3)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACGTCTCATC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (4)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACTCATCTAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (5)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACTCGCGCAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (6)	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGAGCGTAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (7)	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGCGACTAGC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (8)	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGTAGTGATC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (9)	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGTGACACAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (10)	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGTGTATGTC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (11)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATAGATAGAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (12)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATATAGTCGC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (13)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATCTACTGAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (14)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CACGTAGATC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (15)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CACGTGTCGC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (16)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CATACTCTAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (17)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CGACACTATC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (18)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CGAGACGCGC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (19)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CGTATGCGAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (20)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CGTCGATCTC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (21)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTACGACTGC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (22)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTAGTCACTC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (23)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTCTACGCTC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (24)	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTGTACATAC CACAGCGCCATCTGCGCCGA
arrA-CVF1 (25)	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACTAGCAGTA CACAGCGCCATCTGCGCCGA
arrA-CVR1	CCTCTCTATGGGCAGTCGGTGATCCGACGAACTCCYTGYTCCA

**Table S3.** Distribution of the previously described *arrA* gene sequences from the GenBank among different major phylogenetic groups.

Cluster	Accession No.	Identity (%)	Location	Reference
<b>Cluster I</b>	[KR045408, KR045509, KR045417, KR045413, KR045511, KR045421, KR045434, KR045439, KR045437, KR045436, KR045451, KR045432, KR045337, KR045336, KR045396, KR045369, KR045400, KR045406, KR045429, KR045403, KR045405, KR045582, KR045583, KR045521, KR045530, KR045520, KR045418, KR045398, KR045409, KR045420, KR045341, KR045415, KR045412, KR045483, KR045488, KR045502, KR045411, KR045539, KR045534, KR045550, KR045537, KR045515, KR045541, KR045535, KR045312, KR045324, KR045328, KR045318, KR045315, KR04532, KR04532, KR04532, KR045342, KR045331, KR045333, KR045330, KR045404, KR045366, KR045329, KR045419, KR045357, KR045362, KR045345, KR045339, KR045338, KR045372, KR045358, KR045343, KR045410, KR045340, KR045361, KR045367, KR045360, KR045371, KR045359, KR045497, KR045365, KR045416, KR045399, KR045466, KR045498, KR045508, KR045510, KR045514, KR045499, KR045505, KR045540, KR045496, KR045491, KR045478, KR045487, KR045543, KR045489, KR045552, KR045571, KR045518, KR045352, KR045349, KR045492, KR045486, KR045581, KR045522, KR045479, KR045490, KR045559, KR045474, KR045493, KR045531, KR045528, KR045525, KR045458, KR045313, KR045558, KR045568, KR045484, KR045378, KR045517, KR045310, KR045472, KR045501, KR045495, KR045482, KR045580, KR045480, KR045481, KR045565, KR045555, KR045578, KR045579, KR045373, KR045453, KR045468, KR045523, KR045586, KR045566, KR045572, KR045532, KR045473, KR045575, KR045527, KR045475, KR045566, KR045519, KR045467, KR045533, KR045476, KR045463, KR045455]	85-100%	Cambodia (Mekong Delta)	(1)
	KP061061, KP061059, KP061060, KP06107, KP061101, KP061051, KP061138, KP061102, KP061107, KP061039, KP061093, KP061108, KP061089, KP061038, KP061063, KP061064, KP061084, KP061042, KP061070, KP061065, KP061131, KP061053, KP061054, KP061040, KP061097, KP061069, KP061041, KP061114, KP061047, KP061050, KP061048, KP061118, KP061115, KP061117, KP061132, KP061139, KP061126, KP061112, KP061116, KP060963, KP060964, KP060966, KP061035, KP060972, KP060988, KP061001, KP060998, KP061057, KP061134, KP061056, KP061023, KP060974, KP061146, KP061128, KP060975]		Southern China (paddy soil)	(2)
	AB769399, AB769384, AB769388, AB769393, AB769395, AB769389		Japan (paddy soil)	(3)
	JQ041518, JQ0415221, JQ0415222, JQ041520		New Jersey (coastal sediment)	(4)
<b>Cluster II</b>	[KR045447, KR045423, KR045424, KR045426, KR045422, KR045562, KR045569, KR045563, KR045553, KR045560, KR045554, KR045561, KR045379, KR045459, KR045449, KR045375, KR045444, KR045377, KR045391, KR045441, KR045382, KR045464, KR045384, KR045381, KR045394, KR045386, KR045443, KR045374, KR045393, KR045456, KR045452, KR045392, KR045454, KR045442, KR045460, KR045389, KR045356, KR045354]	85-100%	Cambodia (Mekong Delta)	(1)
	[FJ747534, FJ747544, FJ747540, FJ747543, FJ747541, FJ747546, FJ747537, FJ747539, FJ747547, FJ747546, FJ747536, FJ747531, FJ747533, FJ747554, FJ747559, FJ747560, FJ747549, FJ747558, FJ747555, FJ747548, FJ747535, FJ747538]		Chesapeake Bay (sediment)	(5)
	[KP061110, KP061046, KP061140, KP061004, KP061149, KP061150, KP061133, KP061034, KP060994, KP061000, KP060971, KP061121, KP061127, KP061120, KP061123, KP061129, KP061125, KP061119, KP061122]		Southern China (paddy soil)	(2)
	[AB769396, AB769397]		Japan (paddy soil)	(3)
<b>Cluster III</b>	<b><i>Geobacter uraniireducens</i> arrA52 (JF827128), <i>Geobacter</i> sp. OR-1 (AB769875)</b>	85-100%	Southern China (paddy soil)	(2)
	[KP061052, KP061055, KP061049, KP061076, KP061014, KP061075, KP061092, KP061091, KP061018, KP061095, KP061099, KP061073, KP061100]		Chesapeake Bay (sediment)	(5)
	[FJ747538, FJ747535, FJ747548]		Michigan (bioreactor)	(6)
	[JF827135, JF827132, JF827107, JF827134, JF827122, JF827124, F827126, JF827101, JF827105, JF827136, JF827144, JF827103, JF827114, JF827142, JF827143, JF827104, JF827119, JF827145, JF827110, JF827146, JF827138, JF827109, JF827137, JF827127, JF827098, JF827099, JF827128, JF827125, JF827139, JF827141, JF827115, JF827131, JF827100, JF827148, JF827121, JF827129]		Cambodia (sediment)	(7)
	[JQ890224, JQ89022a, JQ89022b, JQ890225]			
<b>Cluster IV</b>	<b><i>Bacillus selenitireducens</i> (AY283639), <i>Bacillus arseniciselenatis</i> (AY660885), <i>Sulfurospirillum barnesii</i> (AY660884), <i>Shewanella</i> (AY660886), <i>Chrysiogenes arsenatis</i> (AY660883), <i>Halarsenatibacter silvermanii</i> (EU723191), <i>Desulfosporosinus</i> (DQ220794),</b>	85-100%	West Bengal (sediment)	(8)
	[FN391000, FN390985, FN391002, FN390991, FN390987, FN390972, FN390998, FN390968, FN390981, FN390973, FN390974, FN390970, FN390992, FN390993, FN390997, FN390969, FN390978, FN390979, FN390958, FN390961, FN390977, FN390955, FN390990, FN390954, FN390983, FN390980, FN390986, FN391003, FN390966, FN390975, FN390984, FN390959, FN390982, FN391005, FN390963, FN390976, FN390962, FN391004, FN390996, FN390994]		Chesapeake Bay (sediment)	(5)
	[FJ747600, FJ747598, FJ747602, FJ747589, FJ747594, FJ747595, FJ747586.]		Cambodia (sediment)	(7)
	[JQ890215, JQ890214, JQ890216, JQ890217, JQ890209, JQ890212, JQ890211, JQ890210, JQ890213, JQ890218, JQ890220, JQ890219, JQ890223]		Japan (paddy soil)	(3)
	[AB769394, AB769386, AB769385, AB769390, AB769400]			
<b>Cluster V</b>	[JF827113, JF827111, JF827147, JF827130, JF827102, JF827140, JF827117, JF827108, JF827120, JF827123, JF827106, JF827118, JF827133.]	85-100%	Michigan (bioreactor)	(6)
	[KR045570, KR045573, KR045585, KR045576, KR045584, KR045401, KR045397, KR045512, KR045513, KR045507, KR045504]			
	[AB769387, AB769391, AB769392]		Cambodia (Mekong Delta)	(1)
	JQ041517, JQ041526, JQ041519		Japan (paddy soil)	(3)
		New Jersey (sediment)	(4)	

**Table S4.** Distribution of the previously described *arrA* gene sequences from the Cache Valley Basin, Utah.

Clusters	<i>ArrA</i> gene clone sequences	Identity (%)	Location
<b>Cluster I</b>	NP8_GW_33 (KF009987.1), NP8_GW_40 (KF009988.1), NP8_GW+G_29, NP8_GW+G_12 (KF009985.1), NP8_GW+G_32, NP3_GW+G_9 (KF009983.1), NP3_GW+G_10 (KF009982.1), NP8_GW+G_16, NP8_GW+G_20 (KF009981.1), NP3_GW+G_29, NP3_GW+G_27 (KF009984.1), NP8_GW+G_45 (NP8_GW+G_27 (KF009986.1), NP1_GW_57 (KF009990.1), NP1_GW_16, NP1_GW_23 (KF009991.1), NP1_GW+G_12, NP3_GW_19, NP3_GW_27 (KF009989.1), NP1_GW_51, NP1_GW_1 (KF009993.1), NP1_GW+G_15 (KF009992.1), NP8_GW+G_33	85-100%	Cache Valley Basin, Utah
<b>Cluster II</b>	NP3_GW_2, NP3_GW+G_20 (KF010068.1), NP3_GW_6 (KF010072.1), NP3_GW_14, NP3_GW_34, NP3_GW+G_17 (KF010077.1), NP1_GW_15, NP3_GW+G_30, NP3_GW_41 (KF010080.1), NP1_GW+G_6 (KF010093.1), NP3_GW_31 (KF010071.1), NP3_GW+G_23 (KF010069.1), NP3_GW_50, NP8_GW+G_22 (KF010078.1), NP3_GW+G_3 (KF010088.1), NP3_GW_8 (KF010073.1), NP3_GW_13, NP3_GW_23, NP3_GW_35 (KF010085.1), NP1_GW+G_4 (KF010076.1), NP3_GW+G_6 (KF010084.1), NP3_GW_26, NP8_GW+G_15, NP3_GW_32, NP3_GW+G_45 (KF010083.1), NP3_GW+G_24 (KF010082.1), NP3_GW+G_8 (KF010091.1), NP8_GW+G_2, NP8_GW+G_46, NP1_GW_17, NP3_GW_1 (KF010086.1), NP3_GW+G_4 (KF010094.1), NP3_GW_28, NP3_GW_30, NP3_GW_11, NP3_GW_10 (KF010081.1), NP3_GW_39, NP3_GW_4 (KF010087.1), NP3_GW+G_36, NP8_GW+G_39, NP3_GW+G_46, NP3_GW+G_15 (KF010090.1), NP3_GW_43 (KF010079.1), NP8_GW+G_47 (KF010074.1), NP3_GW+G_16, NP3_GW_21 (KF010070.1), NP8_GW+G_5 (KF010089.1), NP3_GW_46, NP3_GW_12 (KF010067.1), NP3_GW_40, NP8_GW+G_17 (KF010077.1), NP3_GW_36, NP3_GW_17, NP3_GW_7, NP3_GW_16 (KF010042.1), NP3_GW_20 (KF010043.1), NP3_GW_49 (KF010046.1), NP3_GW_44 (KF010045.1), NP3_GW_22 (KF010047.1), NP3_GW_38, NP3_GW_18, NP8_GW_7 (KF010065.1), NP8_GW_22 (KF010062.1), NP1_GW_4, NP1_GW_35 (KF010049.1), NP1_GW_6, NP3_GW_5, NP3_GW_45 (KF010063.1), NP3_GW_29, NP1_GW+G_8, NP1_GW+G_27 (KF010053.1), NP1_GW_21, NP1_GW+G_43, NP1_GW+G_2 (KF010055.1), NP1_GW+G_50 (KF010058.1), NP1_GW+G_14 (KF010056.1), NP1_GW+G_29, NP1_GW+G_25, NP1_GW+G_9 (KF010060.1), NP1_GW+G_24 (KF010059.1), NP1_GW+G_38 (KF010061.1), NP1_GW+G_13 (KF010057.1), NP1_GW_25, NP1_GW_19 (KF010054.1), NP8_GW+G_41, NP3_GW+G_22, NP8_GW+G_24, NP8_GW+G_23 (KF010038.1), NP8_GW+G_28 (KF010039.1), NP3_GW_15, NP3_GW_37 (KF009996.1), NP8_GW+G_3, NP3_GW+G_37 (KF009997.1), NP8_GW+G_36, NP3_GW+G_2 (KF009998.1), NP1_GW_14 (KF010025.1), NP1_GW_54 (KF010024.1), NP1_GW+G_9 (KF010060.1), NP1_GW_12 (KF010029.1), NP1_GW_55, NP1_GW+G_39, NP1_GW+G_17, NP1_GW+G_34, NP1_GW+G_16, NP1_GW+G_5, NP1_GW+G_23, NP1_GW+G_41 (KF010027.1), NP1_GW+G_44, NP1_GW+G_10 (KF010026.1), NP1_GW+G_7 (KF010030.1), NP1_GW_56, NP1_GW+G_11, NP1_GW+G_33 (KF010033.1), NP1_GW+G_49, NP1_GW+G_40 (KF010032.1), NP1_GW+G_20 (KF010035.1), NP1_GW+G_46 (KF010034.1), NP1_GW_33 (KF010036.1), NP1_GW+G_37 (KF010031.1), NP8_GW+G_31 (KF010037.1), NP8_GW+G_10, NP3_GW+G_38, NP8_GW+G_37, NP8_GW+G_19 (KF009999.1), NP8_GW+G_7, NP8_GW+G_42 (KF010010.1), NP8_GW+G_8 (KF010006.1), NP3_GW+G_1 (KF010012.1), NP3_GW+G_18, NP1_GW+G_47 (KF010009.1), NP3_GW+G_43 (KF010001.1), NP8_GW+G_21 (KF010016.1), NP3_GW+G_34 (KF010021.1), NP8_GW+G_49 (KF010008.1), NP1_GW_26, NP3_GW+G_26, NP3_GW+G_7 (KF010017.1), NP3_GW+G_28 (KF010020.1), NP1_GW+G_21 (KF010002.1), NP3_GW+G_40 (KF010011.1), NP1_GW+G_36, NP3_GW+G_50, NP8_GW+G_4, NP1_GW+G_1, NP8_GW+G_18 (KF010000.1), NP3_GW+G_35 (KF010023.1), NP8_GW+G_35, NP8_GW+G_50 (KF010007.1), NP3_GW+G_12 (KF010019.1), NP1_GW+G_22, NP8_GW+G_48 (KF010018.1), NP3_GW+G_5, NP8_GW+G_13, NP3_GW+G_48, NP3_GW+G_41 (KF010014.1), NP1_GW_53, NP8_GW+G_26 (KF010003.1), NP3_GW+G_33, NP8_GW+G_34 (KF010015.1), NP8_GW+G_43 (KF010005.1), NP3_GW+G_39 (KF010004.1), NP3_GW+G_14 (KF010013.1), NP8_GW+G_38, NP8_GW+G_40, NP1_GW+G_42, NP1_GW_58, NP8_GW+G_14	85-100%	
<b>Cluster III</b>	NP8_GW_29 (KF009949.1), NP3_GW_48 (KF009950.1), NP3_GW_47 (KF009948.1), NP8_GW_23, NP8_GW_5 (KF009963.1), NP8_GW_34 (KF009968.1), NP8_GW_41, NP8_GW_2 (KF009949.1), NP8_GW_30 (KF009972.1), NP8_GW_6 (KF009960.1), NP8_GW_21, NP8_GW_12, NP8_GW_3 (KF009958.1), NP8_GW_49, NP8_GW_26, NP8_GW_11 (KF009978.1), NP8_GW_35 (KF009967.1), NP8_GW_14 (KF009962.1), NP8_GW_38, NP8_GW_36, NP8_GW_9 (KF009974.1), NP8_GW_45 (KF009969.1), NP8_GW_28, NP8_GW_43, NP8_GW_10, NP8_GW_44 (KF009971.1), NP8_GW_25 (KF009966.1), NP8_GW_46 (KF009980.1), NP8_GW_50 (KF009965.1), NP8_GW_27, NP8_GW_39 (KF009961.1), NP8_GW_31 (KF009958.1), NP1_GW+G_48 (KF010064.1), NP1_GW+G_28 (KF010075.1), NP3_GW+G_25, NP3_GW+G_13 (KF010092.1), NP8_GW+G_25 (KF010066.1), NP3_GW_24, NP1_GW_30 (KF010052.1), NP1_GW+G_3, NP1_GW_39 (KF010051.1), NP1_GW_11 (KF010050.1), NP3_GW_25, NP3_GW+G_42, NP3_GW_33 (KF010048.1), NP8_GW_48 (KF009970.1)	85-100%	
<b>Cluster IV</b>	NP8_GW_15 (KF009955.1), NP1_GW_5 (KF009947.1), NP8_GW_32 (KF009977.1), NP8_GW_27, NP3_GW_9 (KF009952.1), NP3_GW_3 (KF009951.1), NP1_GW_37 (KF009946.1), NP1_GW_24, NP1_GW_27 (KF009945.1), NP3_GW_42	85-100%	
<b>Cluster V</b>	NP1_GW_34 (KF010109), NP1_GW_3 (KF010096), NP1_GW_52NP1_GW_63 (KF010101), NP1_GW_67, NP1_GW_60 (KF010103), NP1_GW_66, NP1_GW_65, NP1_GW_40, NP1_GW_9 (KF010107), NP1_GW_8 (KF010106), NP1_GW_32 (KF010108), NP1_GW_62 (KF010102), NP1_GW+G_32 (NP1_GW_7, NP1_GW+G_45 (KF010105), NP1_GW_64 (KF010100), NP1_GW_59 (NP1_GW_61 (KF010110), NP1_GW_13 (KF010095), NP1_GW_36 (KF010097), NP1_GW_29 (NP8_GW+G_1 (NP3_GW+G_19 (NP8_GW+G_9 (KF010112), NP8_GW+G_11 (KF010111), NP3_GW+G_31 (KF010114), NP8_GW+G_44 (KF010113), NP8_GW+G_30 (NP1_GW+G_26 (KF010104), NP1_GW+G_30 (KF010098), NP1_GW+G_18 (KF010099), NP1_GW+G_31, NP1_GW_38	85-100%	