Lateral Gene Transfer in a Heavy Metal-Contaminated-Groundwater Microbial Community

Christopher L. Hemme  
*University of Rhode Island, hemmecl@uri.edu*

Stefan J. Green

*See next page for additional authors*

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

Follow this and additional works at: [https://digitalcommons.uri.edu/bps_facpubs](https://digitalcommons.uri.edu/bps_facpubs)

Citation/Publisher Attribution

Available at: [http://dx.doi.org/10.1128/mBio.02234-15](http://dx.doi.org/10.1128/mBio.02234-15)

This Article is brought to you for free and open access by the Biomedical and Pharmaceutical Sciences at DigitalCommons@URI. It has been accepted for inclusion in Biomedical and Pharmaceutical Sciences Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons@etal.uri.edu](mailto:digitalcommons@etal.uri.edu).
Authors
Christopher L. Hemme, Stefan J. Green, Lavanya Rishishwar, Om Prakash, Angelica Pettenato, Romy Chakraborty, Adam M. Deutschbauer, Joy D. Van Nostrand, Liyou Wu, Zhili He, I. King Jordan, Terry C. Hazen, Adam P. Arkin, Joel E. Kostka, and Jizhong Zhou

This article is available at DigitalCommons@URI: https://digitalcommons.uri.edu/bps_facpubs/84
Lateral Gene Transfer in a Heavy Metal-Contaminated-Groundwater Microbial Community

Christopher L. Hemme, Stefan J. Green, Lavanya Rishishwar, Om Prakash, Angelica Pettenato, Romy Chakraborty, Adam M. Deutschbauer, Joy D. Van Nostrand, Liyou Wu, Zhili He, I. King Jordan, Terry C. Hazen, Adam P. Arkin, Joel E. Kostka, Jizhong Zhou

ABSTRACT Unraveling the drivers controlling the response and adaptation of biological communities to environmental change, especially anthropogenic activities, is a central but poorly understood issue in ecology and evolution. Comparative genomics studies suggest that lateral gene transfer (LGT) is a major force driving microbial genome evolution, but its role in the evolution of microbial communities remains elusive. To delineate the importance of LGT in mediating the response of a groundwater microbial community to heavy metal contamination, representative *Rhodanobacter* reference genomes were sequenced and compared to shotgun metagenome sequences. 16S rRNA gene-based amplicon sequence analysis indicated that *Rhodanobacter* populations were highly abundant in contaminated wells with low pHs and high levels of nitrate and heavy metals but remained rare in the uncontaminated wells. Sequence comparisons revealed that multiple geochemically important genes, including genes encoding Fe(II)/Pb(II) permeases, most denitrification enzymes, and cytochrome c553, were native to *Rhodanobacter* and not subjected to LGT. In contrast, the *Rhodanobacter* pangenome contained a recombinational hot spot in which numerous metal resistance genes were subjected to LGT and/or duplication. In particular, Co(II)/Zn(II)/Cd(II) efflux and mercuric resistance operon genes appeared to be highly mobile within *Rhodanobacter* populations. Evidence of multiple duplications of a mercuric resistance operon common to most *Rhodanobacter* strains was also observed. Collectively, our analyses indicated the importance of LGT during the evolution of groundwater microbial communities in response to heavy metal contamination, and a conceptual model was developed to display such adaptive evolutionary processes for explaining the extreme dominance of *Rhodanobacter* populations in the contaminated groundwater microbiome.

IMPORTANCE Lateral gene transfer (LGT), along with positive selection and gene duplication, are the three main mechanisms that drive adaptive evolution of microbial genomes and communities, but their relative importance is unclear. Some recent studies suggested that LGT is a major adaptive mechanism for microbial populations in response to changing environments, and hence, it could also be critical in shaping microbial community structure. However, direct evidence of LGT and its rates in extant natural microbial communities in response to changing environments is still lacking. Our results presented in this study provide explicit evidence that LGT played a crucial role in driving the evolution of a groundwater microbial community in response to extreme heavy metal contamination. It appears that acquisition of genes critical for survival, growth, and reproduction via LGT is the most rapid and effective way to enable microorganisms and associated microbial communities to quickly adapt to abrupt harsh environmental stresses.

A abrupt, intense environmental stress can overwhelm the ability of microbial communities to adapt by classical selection/ drift/mutation mechanisms and force communities to rapidly adapt to such a stress by large-scale genomic rearrangements, including lateral gene transfer (LGT) and gene duplication (GD) (1–12). These genomic rearrangement events frequently underlie...
the natural attenuation of environmental pollutants, leading to important ecological and economic consequences (13). However, it remains challenging to identify and quantify such events from metagenome data alone (1, 6, 12). Most metagenomes are too diverse and complex to allow for extensive assembly, making it difficult to distinguish between native and laterally transferred genes. To ameliorate this problem, sequencing of the genomes of isolated microbes from the environments in question can serve as references for comparison (14–17). This strategy has been extensively used in the Human Microbiome Project to identify pathways of gene sharing between community members and hosts (10, 14, 17–23). However, for many types of environmental samples, such as soil and subsurface groundwater, few reference genomes are available, and thus, a large proportion of such communities are unrepresented in genomic databases (1).

The Oak Ridge Integrated Field Research Challenge (OR-IFRC) site at the Y-12 Federal Security Complex in Oak Ridge, TN, is a well-characterized experimental field site for studying the environmental impacts of legacy waste on soil and groundwater systems (1, 24–27). The groundwater contamination plume at this site is derived from spent uranium and nitric acid waste which was stored in unlined ponds that were capped in 1983. Near-source zone groundwater contains chronically high concentrations of radionuclides (e.g., uranium), nitric acid, organics, salts, mercury, and other heavy metals, resulting in an extremely low diversity of subsurface microbial communities (1, 27, 28). Repeated cultivation-independent analyses of community genomic DNA and RNA from groundwater recovered from well FW106 within the near-source zone revealed a stressed microbial community of low diversity, dominated by populations of metal-resistant, denitrifying *Rhodanobacter* (>80% by analysis of rRNA gene abundance) (1, 24, 29–33). The FW106 metagenome showed high relative abundances of genes encoding geochemical resistance functions required for microbial survival in the presence of known environmental contaminants at the site (1), including those associated with denitrification, heavy metal resistance, and organic compound degradation. These genes were more abundant in the FW106 metagenome than in a control groundwater community from well FW301 at the OR-IFRC background site (1, 34). Experimental and metagenomic analyses have shown that LGT of heavy metal resistance genes may partially account for their high abundance in the FW106 metagenome (1). To complement metagenomic analyses, multiple laboratories are involved in isolating and sequencing of reference genomes from these environments, including a number of strains of dominant *Rhodanobacter* populations (24, 32, 35–39). These reference genomes are valuable for delineation of the importance of LGT in driving the adaptation of the groundwater microbial communities in response to extremely heavy environmental contamination.

Overabundance of resistance systems, as revealed by several previous studies, can result from selection favoring naturally resistant populations (30, 40–42), via LGT, and/or via GD (1, 6, 7, 10, 11, 43, 44). Three hypotheses can provide explanations for the overabundance of geochemical resistance genes in the contaminant-stressed community: (i) they provide no selective benefit in the contaminated environment and they are overabundant simply because the host organisms are abundant, (ii) they provide selective benefit to the hosts under contaminated conditions but are not subjected to LGT or GD, and (iii) they provide selective benefit to the host under contaminated conditions and are subjected to LGT and/or GD. To test these hypotheses, multiple *Rhodanobacter* species were isolated from OR-IFRC sites, and their genomes were sequenced and compared to those of other *Rhodanobacter* reference strains and to OR-IFRC site groundwater metagenomes. The metagenomes and several of the *Rhodanobacter* strains were sequenced as part of previous studies, and seven additional *Rhodanobacter* strains were sequenced as part of this study to complement this data set. Comparative analysis showed that many geochemically important genes were highly abundant in the uranium-contaminated groundwater community due to the dominance of *Rhodanobacter* in the environment. A recombinational hot spot was identified in the *Rhodanobacter* pangenome where multiple metal resistance genes were acquired via LGT and/or GD. These results indicated that LGT could play a critical role in driving the evolution of a groundwater microbial community in response to extreme heavy metal contamination.

**RESULTS**

**Overabundance of *Rhodanobacter* species in contaminated wells.** A 16S rRNA gene survey of groundwater microbial communities from 100 wells located within the OR-IFRC site was conducted in the years 2012 to 2013 to identify the microbial community composition across the OR-IFRC research area (45). To identify both the locations and the relative abundances of *Rhodanobacter* strains in these communities, 16S rRNA gene sequences from *Rhodanobacter* isolates were obtained from the Ribosomal Database (46) and compared to the 16S rRNA gene sequences from the OR-IFRC site 100-well survey by BLASTn (47). A total of 370,765 OR-IFRC sequences showed hits to *Rhodanobacter* sequences with >97% sequence identity across the aligned regions (see Table S1 in the supplemental material). Of these sequences, 79.8% were mapped to the three wells located in contaminated areas 1 and 3, with 33.8% to FW106 (area 3), 27.4% to FW109 (area 3), and 18.6% to FW021 (area 1) (see Table S1). These wells all lie near the base of contaminated groundwater plumes emanating from the former S-3 waste disposal ponds and are heavily contaminated with uranium, nitric acid, and heavy metals. In contrast to the contaminated sites, only a small number of hits were seen for background (i.e., uncontaminated) wells (e.g., FW300, FW301, and FW303), with <0.1% for each (see Table S1). Thus, while *Rhodanobacter* species are broadly distributed across the OR-IFRC research area, they are most dominant in low-pH groundwater located near the source of the contaminant plumes.

**Abundance of geochemically important genes in the FW106 metagenome compared to *Rhodanobacter* species.** Previous comparisons of OR-IFRC contaminated-groundwater metagenomes indicated that numerous geochemically important genes were abundant in the contaminated-groundwater metagenome and that some of the abundant genes could have been acquired by LGT (1, 34). However, LGT events are difficult to distinguish with metagenomic sequences alone. To determine whether these overabundant genes were linked to *Rhodanobacter* populations in the contaminated groundwater community and whether LGT contributed to the abundance of these genes, metagenomes from *Rhodanobacter*-dominated uranium-contaminated groundwater communities (OR-IFRC well FW106) and background groundwater communities (OR-IFRC well FW301) were compared directly to the *Rhodanobacter* reference genomes at the species and genus levels (Table 1 and Fig. 1; see also Table S2 in the supplemental material). Previous comparative analysis of the two met-
agenomes showed that the abundance of *Rhodanobacter* populations was <1% in the background groundwater metagenome compared to >85% in the contaminated groundwater metagenome (30). The abundance of each cluster of orthologous groups (COG) of interest was calculated for each sample and for all sequenced bacterial genomes available as of September 2014, and this information was used to calculate ln odds ratios for each COG (Fig. 1). COGs of interest were divided into four categories: nitrate reduction, Fe/Pb permease, and cytochrome (H+/H2). COGs of interest were divided into four categories: nitrate reduction, Fe/Pb permease, and cytochrome (H+/H2).

All *Rhodanobacter* strains studied showed high abundances of genes related to denitrification, divalent metal ion efflux (czc), Fe/Pb permease, and cytochrome oxo, similar to the contaminated groundwater metagenome (Fig. 1). While the abundance profile of an isolate is unlikely to exactly match that of even a simple metagenome, the abundance profiles of geochemically important genes (including genes encoding denitrification enzymes, hydrogenases, and some cytochromes) in several *Rhodanobacter* strains resembled those of the contaminated groundwater metagenome (Fig. 1). The abundance profiles of the contaminated-groundwater metagenome most closely mirrored those of OR-IFRC isolates *Rhodanobacter* strains 2APBS1, *Rhodanobacter* species strains OR87, OR92, OR444, FW510-R12, FW104-R3, and less closely resembled OR-IFRC isolates *Rhodanobacter* sp. strain 115 and *Rhodanobacter* strains 116-2 and the three non-OR-IFRC isolates (Fig. 1). The hydrogenase gene profiles showed that the contaminated-groundwater metagenome was most similar to the genomes of *Rhodanobacter* strains 2APBS1 and 116-2, *Rhodanobacter* species strains OR87, OR444, FW104-R8, FW510-R10, FW510-R12, and FW510-T8, and *Rhodanobacter thiooxydans* (Fig. 1). These results show that while the contaminated-groundwater metagenome does resemble the genomes of the dominant *Rhodanobacter* populations, the *Rhodanobacter* populations cannot fully account for all of the metabolic potential in the contaminated-groundwater metagenome.

**Identification of LGT events in *Rhodanobacter* genomes.** Some of the overabundance of genes in the contaminated-groundwater metagenome could be attributed to genes naturally present in the dominant *Rhodanobacter* populations. However, resistance genes could also be acquired from other species through LGT. To identify likely causes of resistance gene overabundance in the contaminated-groundwater metagenome, the *Rhodanobacter* genomes were searched for target genes overabundant in the contaminated-groundwater metagenome. Putative LGT events in the genomes were then identified and correlated to the genes of interest. Multiple methods to identify aberrant sequences deviating from the genomic background were used to identify LGT events in the isolate genomes, including phylogenetic analysis, DarkHorse (48), Alien_Hunter (49), and GOHTAM (50, 51) (Fig. 2; see also Fig. S1 and S2 and Table S2 in the supplemental material). Analysis of the complete genome of *R. denitrificans* 2APBS1 revealed several regions containing putative LGT events, many of which matched up to annotated transposase and integrase genes (Fig. 2). Fifty-one genes related to heavy metal resistance were also analyzed (see Table S2). Of these genes, seven were identified by three LGT detection programs as putative LGT events, 13 were detected by at least two programs, 19 were detected by a single program, and 12 were not identified as LGT events by any of the programs used (see Table S2). In particular, most of the genes in adjacent arsenate resistance (ars) and copper resistance (cop) operons were identified as LGT events by all programs used (see Table S2). Many of these genes were located within a large genomic region near the 1.7-to-2.3-Mb mark noted for a large number of metal resistance genes, including operons encoding mercuric, arsenate, and divalent cation resistance (e.g., cze) (Fig. 2). DarkHorse analysis indicated the sources of these genes to be multiple gamma- and betaproteobacteria known to be present at OR-IFRC sites, including *Burkholderia*, *Herbaspirillum*, *Ralstonia*, *Cupriavidus*, *Pseudomonas*, and *Stenotrophomonas* (see Table S2). In the following sections, detailed descriptions of gene abundance and possible LGT events for several geochemically important genes are provided.

**Nitrate/nitrite antiporter gene abundance.** While most denitrification genes were present in high abundance in the FW106

### Table 1 Genomes and metagenomes used in this study

<table>
<thead>
<tr>
<th>Source of genome or metagenome</th>
<th>DSM no.</th>
<th>OR-IFRC strain</th>
<th>Denitrification from nitrate*</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rhodanobacter isolate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>R. denitrificans</em> 2APBS1T</td>
<td>DSM23569</td>
<td>Y</td>
<td>Y</td>
<td>31</td>
</tr>
<tr>
<td><em>R. denitrificans</em> 116-2</td>
<td>DSM24678</td>
<td>Y</td>
<td>Y</td>
<td>31</td>
</tr>
<tr>
<td>Sp. 115</td>
<td>Y</td>
<td>N</td>
<td>?</td>
<td>24</td>
</tr>
<tr>
<td>Sp. OR87</td>
<td>Y</td>
<td>?</td>
<td>?</td>
<td>24</td>
</tr>
<tr>
<td>Sp. OR92</td>
<td>Y</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Sp. OR444</td>
<td>Y</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Sp. FW104-R3</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>This study</td>
</tr>
<tr>
<td>Sp. FW104-R5</td>
<td>Y</td>
<td>?</td>
<td>?</td>
<td>This study</td>
</tr>
<tr>
<td>Sp. FW104-R8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>This study</td>
</tr>
<tr>
<td>Sp. FW104-T7</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>This study</td>
</tr>
<tr>
<td>Sp. FW510-T8</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>This study</td>
</tr>
<tr>
<td>Sp. FW510-R10</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>This study</td>
</tr>
<tr>
<td>Sp. FW510-R12</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>This study</td>
</tr>
<tr>
<td><em>R. thiooxydans</em></td>
<td>DSM18863</td>
<td>N</td>
<td>Y</td>
<td>31</td>
</tr>
<tr>
<td><em>R. fulvis</em></td>
<td>DSM18449</td>
<td>N</td>
<td>N</td>
<td>31</td>
</tr>
<tr>
<td><em>R. spathiphylli</em></td>
<td>DSM17631</td>
<td>N</td>
<td>N</td>
<td>31</td>
</tr>
<tr>
<td><strong>OR-IFRC metagenomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FW106</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>FW301</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>34</td>
</tr>
</tbody>
</table>

*Indicates presence or absence of denitrification from nitrate based on experimental and/or genomic evidence. Denitrification from nitrite may still be possible in strains incapable of nitrate reduction.
metagenome, the putative nitrate/nitrite antiporter (COG2223) genes showed particularly high abundance (Table 1 and Fig. 1A) (1). The numbers of putative nitrate/nitrite antiporter genes varied in the *Rhodanobacter* strains, with 10 strains (*R. denitrificans* 2APBS1, *R. denitrificans* 116-2, *Rhodanobacter* species OR87, OR92, OR444, FW104-R3, FW104-R5, FW104-T7, and FW510-R12, and *R. thiooxydans*) containing genes encoding a pair of nitrate/nitrite transporters associated with denitrification pathways (represented by homologs R2APBS1_2892 and R2APBS1_2893 from *R. denitrificans* 2APB1) (Fig. 3).

Phylogenetic analysis of the nitrate/nitrite transporter (COG2223) genes provided new insights into the population structure of the uranium-contaminated community. Homologs of *R. denitrificans* 2APBS1_2892 were separated into two clades (Fig. 3; see also Fig. S3 in the supplemental material). Ten homologs were found in clade 1a and represented the vertically inherited ancestral *Rhodanobacter* genes (Fig. 3). Three strains (*Rhodanobacter* species strains FW104-R8, FW510-T8, and FW510-R10) lacked this particular suite of genes but had a similar group of genes in a different orientation, appearing to be derived
from species other than *Rhodanobacter* (represented by clade 1b in Fig. 3). Genes in this cluster were more similar to genes from *Thermomonas, Stenotrophomonas*, and *Pseudomonas* species than from *Rhodanobacter* and were associated with transposable elements. Clade 2a (represented by R2APBS1_2893) contains 14 FW106 homologs (Fig. 3). *Rhodanobacter* sp. 115 did not have a clade 2a homolog, and *Rhodanobacter fulvus* contained a distinct homolog representing its own clade (clade 2b) (Fig. 3). A third putative nitrate/nitrite transporter gene (clade 4a) was not associated with denitrification genes and was found in all 14 *Rhodanobacter* strains. The region surrounding this gene appeared to be native to *Rhodanobacter*, but the gene itself was most similar to *Burkholderia* homologs. Clade 5 (represented by R2APBS1_2043 from *R. denitrificans* 2APB1) was found in 11 OR-IFRC strains and was absent in the three non-OR-IFRC reference genomes (Fig. 3). This homolog was not associated with denitrification genes but with a cluster of $\text{Cd}^{2+}/\text{Zn}^{2+}/\text{Co}^{2+}$ (*czc*) efflux genes located between a pair of transposons. While the genealogy of the region as a whole was unclear, several of the genes were most similar to betaproteobacterial homologs.
Previous analyses of the contaminated-groundwater metagenome suggested the existence of multiple *Rhodanobacter* populations (1). The *Rhodanobacter* COG2223 amino acid sequences were compared to homologs from the contaminated- and background groundwater metagenomes to identify possible population structures in these metagenomes (see Fig. S4A in the supplemental material). A partial clade 1a nitrate/nitrite transporter gene sequence from contaminated groundwater was partially reconstructed from four contiguous gene fragments (FW106_100021117–FW106_100021119) on scaffold FW106_c1000211.
(19.8 kb), which was most similar to the *R. thiooxydans* homolog (see Fig. S3A). The R2APBS1_2982 homologs in the contaminated-groundwater metagenome homologs formed a distinct subclade within the larger *Rhodanobacter* clade. A similar pattern was observed for the contiguous R2APBS1_2893 homolog, which matched two contaminated-groundwater and one background groundwater metagenome fragment that showed similarity to the genome of *R. thiooxydans* (see Fig. S3B). This evidence supports the hypothesis that multiple *Rhodanobacter* populations exist in the contaminated-groundwater community.

**Metal transport and metabolism gene abundance.** A variety of metal resistance genes were identified in high abundance in the contaminated-groundwater metagenome, including *mer* operon genes, high-affinity Fe\(^{2+}/\text{Pb}^{2+}\) permease genes, and Cd\(^{2+}/\text{Zn}^{2+}/\text{Co}^{2+}\) (czc) efflux genes, for example (Fig. 1B) (1). Most of these genes were also present in *Rhodanobacter* genomes at variable abundances (Fig. 1B). *cczA* was the most abundant gene, especially in the contaminated-groundwater metagenome and the genomes of *R. denitrificans* 2APBS1, *R. thiooxydans* 116-2, *R. thiooxydans*, and *Rhodanobacter* species OR87, OR92, OR444, FW104_R3, FW104_R5, FW104_T7, and FW510_R12. *Rhodanobacter* sp. 115 and *R. spathiphylli* had lower abundances of *cczD* than other isolates, and FW106 showed a lower abundance of *cczB* than other isolates (Fig. 1B).

Two copies of high-affinity Fe\(^{2+}/\text{Pb}^{2+}\) permease genes were identified from *Rhodanobacter* genomes (Fig. 1B). The first copy, represented in *R. denitrificans* 2APBS1 by coding sequence (CDS) R2APBS1_2568, was present in all strains except *Rhodanobacter* sp. 115, where this region appeared to have been excised. This region was highly conserved in several genomes (*Rhodanobacter* sp. 115 being the primary exception), and no transposable elements were identified. The second copy, represented in *R. denitrificans* 2APBS1 by CDS R2APBS1_1742, lay in the high-LGT region described above that was identified by DarkHorse analysis. All *Rhodanobacter* species contained a copy of this gene except for *R. spathiphylli* and *Rhodanobacter* sp. 115, and the latter may have been rendered a pseudogene due to a frameshift mutation. As previously mentioned, this region was associated with other metal resistance genes, including *Cu^{2+}\) efflux, *Czc\) efflux, arsenate resistance, and iron efflux genes. *Rhodanobacter* sp. 115 contains an Fe/Pb permease gene that is distinct from other *Rhodanobacter* homologs (Fig. 1B).

Mercury is a major contaminant at the OR-IFRC, and existing experimental evidence has suggested that mercury resistance genes were subjected to LGT in this environment (8). Mercuric resistance genes were present in high abundance in both the contaminated- and background groundwater metagenomes across a wide phylogenetic breadth, and multiple mercuric resistance operons were identified in *Rhodanobacter* genomes (see Fig. S4 to S6 in the supplemental material). One mercuric resistance operon (represented by *merA* gene R2APBS1_1852 of *R. denitrificans* 2APBS1) appeared to be common to *R. denitrificans* 2APBS1, *Rhodanobacter* sp. OR87, *Rhodanobacter* sp. OR92 (all OR-IFRC strains), *R. thiooxydans*, and possibly, OR-IFRC strain *Rhodanobacter* sp. 115. This operon appeared to have been duplicated twice in *R. denitrificans* 2APBS1 (*merA* genes R2APBS1_2075 and R2APBS1_2097) (see Fig. S5 and S6).

**Hydrogenase and cytochrome gene abundance.** The abundances of hydrogenase systems varied widely between *Rhodanobacter* strains (Fig. 1C). Based on the abundance profiles of hydro-
R. denitrificans, Rhodanobacter sp. 115, and R. thio- 
oxidans also contained an ars arsenic resistance operon (Fig. 5; see also Fig. S7). Multiple instances of czc operon genes, copper resistance genes (copAB and cueR), lead resistance genes (pbrA), and uncharacterized heavy metal resistance genes were found throughout the genomes, often in lineage-specific arrangements (Fig. 4; see also Fig. S7). Of the strains isolated from wells FW104 and FW510, FW104_R5 closely resembled R. denitrificans 2APBS1 and most of the other strains resembled Rhodanobacter species OR87 (Fig. 4). Only R. denitrificans 2APBS1, R. thiooxydans, and Rhodanobacter species 115 and FW104_R5 contained arsenate reductase genes while containing the conserved mercuric reductase operon (Fig. 4).

**DISCUSSION**

The mechanism by which specific bacterial populations adapt to intensive stress resulting from multiple environmental contaminants and extreme conditions remains an open question. Several experimental analyses at the OF-IFRC sites suggest that nitrate concentration, pHs, and heavy metal concentrations are the key drivers of community structure and functions in contaminated sites (30, 41), and many strains have been isolated from these sites that show resistance to these particular environmental stresses (28, 32, 37, 39, 52, 53). Comparative genomics analyses of the Rhodanobacter genomes and the OR-IFRC groundwater metagenomes show that many of the genes identified as overabundant in the contaminated-groundwater metagenome are present in most of the Rhodanobacter strains. Some of these genes, such as the core denitrification pathway genes, are vertically inherited, with no evidence of lateral gene transfer. Many metal resistance genes, including czc and mercuric resistance genes, show evidence of extensive lateral gene transfer and/or gene duplication in Rhodanobacter strains. An area of the Rhodanobacter pangenome appears to be a recombinational hot spot that allows rapid acquisition and accumulation of these genes in Rhodanobacter populations. Thus, the overabundance of putative resistance genes in the contaminated-groundwater population can be attributed to a combination of vertical inheritance, LGT, and GD in the dominant Rhodanobacter population (Table 2).

However, it should be noted that the methods used for identifying putative LGT events all use different strategies to detect regions of the genome that deviate from the genomic background. Of the 51 putative laterally transferred genes listed in Table S2 in the supplemental material, 76.5% were identified as positive LGT events by at least one of the programs used (Alien_Hunter, Dark-Horse, and GOHTAM) and 39.2% were identified as positive LGT events by at least two methods. Many of these genes are located in operons, and thus, it can be assumed that if some of the genes in the cluster were identified as laterally transferred, then all genes in the cluster were likely laterally transferred as well. Additional evidence supporting lateral transfer of these genes is their phylogenetic similarity to orthologs from organisms other than Rhodanobacter and the association of many of these genes in the Rhodanobacter genomes with mobile elements.

Based on the results of this work, a conceptual model was constructed to predict how Rhodanobacter populations adapted to contamination and came to dominance in many of the OR-IFRC contaminated-groundwater communities. Comparative genom-
A) **Pristine community**
- High biodiversity
- Even distribution
- High natural abundance of geochemical resistance genes (e.g. czc)

B) **Stressed community (exposure)**
- Introduction of high chronic concentrations of contaminants (e.g. nitrate, heavy metals) and reduction of pH
- Loss of species diversity
- Dominance by naturally resistant populations (*Rhodanobacter*, blue)

C) **Stressed community (adapted)**
- Acquisition of laterally transferred resistance genes (e.g. czc)
- High diversity of metal resistance genes in recombinational hotspot of *Rhodanobacter* genomes (red dot)
- Duplication of some resistance genes (e.g. mer)

**FIG 5** A conceptual model of *Rhodanobacter* population dynamics in contaminated-groundwater systems. Populations are distinguished by different colors, with *Rhodanobacter* shown in blue. Fuzzy red dots indicate the proposed recombinational hot spot in the *Rhodanobacter* pan-genome. (A) Background state with high biodiversity and low abundance of *Rhodanobacter*. (B) Contaminated state with chronic exposure to uranium, nitrate, and low pH in which the environment selects for contamination. (C) Lateral transfer and duplication of geochemical resistance genes (indicated by arrows) within and between *Rhodanobacter* populations, which initiates subsequent rounds of selection.

**TABLE 2** Lateral gene transfer and gene duplication events in *Rhodanobacter* species relevant to geochemical resistance in the FW106 contaminated-groundwater metagenome

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Eventa</th>
<th>Stressor</th>
<th>Gene or product</th>
<th>Representative CDS(s)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. No selection in response to presence of stressors, no LGT/GD</td>
<td>VI</td>
<td>None</td>
<td>Housekeeping genes</td>
<td></td>
</tr>
<tr>
<td>IIa. Selection in response to presence of stressors, no LGT/GD</td>
<td>VI</td>
<td>Heavy metals</td>
<td>High-affinity Fe(^{2+})/Pb(^{2+}) permease</td>
<td>R2APBS1_2568</td>
</tr>
<tr>
<td></td>
<td>VI</td>
<td>Heavy metals</td>
<td>Cytochrome c(_{553})</td>
<td>R2APBS1_0257</td>
</tr>
<tr>
<td>IIb. Selection in response to lack of stressors, no LGT/GD</td>
<td>GL</td>
<td>Nitrate/nitrite</td>
<td>Nitrate/nitrite transporter</td>
<td>R2APBS1_2892</td>
</tr>
<tr>
<td>IIIa. Selection in response to presence of stressors, LGT/GD predating contamination</td>
<td>LGT</td>
<td>Heavy metals</td>
<td>High-affinity Fe(^{2+})/Pb(^{2+}) permease</td>
<td>R2APBS1_1742</td>
</tr>
<tr>
<td></td>
<td>LGT</td>
<td>Heavy metals (mercury)</td>
<td>Mercuric resistance operon genes (original)</td>
<td>R2APBS1_1850–1856, c R2APBS1_2076–2071, c R2APBS1_2095–2101c</td>
</tr>
<tr>
<td>IIIb. Selection in response to presence of stressors, LGT/GD following contamination</td>
<td>GD</td>
<td>Heavy metals (mercury)</td>
<td>Mercuric resistance operon genes (duplicates)</td>
<td>R2APBS1_1850–1856, c R2APBS1_2076–2071, c R2APBS1_2095–2101c</td>
</tr>
<tr>
<td></td>
<td>LGT</td>
<td>Heavy metals (divalent cations)</td>
<td>Co(^{2+})/Zn(^{2+})/Cd(^{2+}) efflux genes</td>
<td>R2APBS1_1735–1751</td>
</tr>
<tr>
<td></td>
<td>LGT</td>
<td>Heavy metals (arsenic)</td>
<td>Arsenate reductase operon genes</td>
<td>R2APBS1_1729–1733</td>
</tr>
<tr>
<td></td>
<td>LGT</td>
<td>Nitrate/nitrite</td>
<td>Clade 2b nitrate/nitrite transporters</td>
<td>RhoFW510R12_08626</td>
</tr>
<tr>
<td>IV. Abundant genes that cannot currently be characterized</td>
<td>Unknown</td>
<td>NiFe hydrogenase III (large and small subunits)</td>
<td>R2APBS1_2173–2174</td>
<td></td>
</tr>
</tbody>
</table>

a VI, vertical inheritance; LGT, lateral gene transfer; GD, gene duplication; GL, gene loss.
b The locus tags of representatives of the gene families in the reference genome *R. denitrificans* 2APBS1 are listed.
c It is unclear which is original and which are duplicates.
suggests that viral and/or plasmid vectors are the primary means of LGT. It remains to be seen if plasmids specific to *Rhodanobacter* exist or if *Rhodanobacter* species are capable of conjugation. It appears that *Rhodanobacter* may be a key node in such a network in contaminated-groundwater environments.

A second interesting question regards whether the observed LGT/GD events preceded contamination of the site or whether these are recent changes specifically in response to contamination. Some metal resistance genes, for instance, are conserved in most of the strains, suggesting they were acquired prior to contamination. On the other hand, the OR-IFRC strains often show genomic properties distinct from those of the reference strains isolated outside the OR-IFRC sites, particularly in terms of their complement of metal resistance genes. While population structures are evident in the OR-IFRC strains, it is still difficult to determine whether these represent legacy (fixed in the population prior to contamination) or opportunistic (fixed in the population following contamination) events. While definitive evidence of lateral gene transfer is difficult to identify by using sequence data alone, the biogeography and population structures of *Rhodanobacter* within and outside the OR-IFRC show compelling evidence of contamination-driven lateral gene transfer and gene duplication. In particular, the apparent duplication of mercuric resistance operons in *R. denitrificans* species provides the best evidence to date of a genome-scale evolutionary event specifically in response to contamination.

Another interesting question is whether the proposed model of *Rhodanobacter* may be extended to other bacterial populations at the OR-IFRC sites. For instance, do other naturally resistant OR-IFRC populations, such as *Pseudomonas* and *Stenotrophomonas*, exhibit the same selection-LGT selection dynamics as *Rhodanobacter*? If so, do these populations employ the same mechanisms for accumulating these genes? Analysis of the OR-IFRC background groundwater metagenome from the OR-IFRC background site suggests a reservoir of mobile geochemical resistance genes in groundwater communities (34). If multiple populations in the contaminated-groundwater source zone are drawing from the same pool of mobile genes, different OR-IFRC populations might show similar LGT patterns. Alternatively, individual population genome dynamics may affect the rate of gene transfer and fixation in specific populations, which in turn might affect their relative fitness in these environments. As more OR-IFRC genomes become available, these hypotheses can be tested directly.

The results of this analysis help explain why *Rhodanobacter* species are dominant in many OR-IFRC contaminated environments, but some questions remain. It is known, for instance, that *Rhodanobacter* populations exist in uncontaminated background communities, albeit in extremely low abundances. It is unknown whether the introduction of contaminants allows opportunistic *Rhodanobacter* populations to grow to dominance or whether these environments are colonized by exogenous *Rhodanobacter* populations. It is also not clear why *Rhodanobacter* populations dominate certain contaminated environments and not others. That is, do specific environments select specific populations, or are the dominance patterns observed at the OR-IFRC the result of stochastic effects (54–57)? Further study of this phenomenon may help explain why other, seemingly equally adapted populations are not able to outcompete *Rhodanobacter* at specific sites and vice versa.

In conclusion, comparative genome analysis of *Rhodanobacter* isolates and metagenomes from contaminated-groundwater sites clarifies the role and extent of LGT and GD in the accumulation of geochemical resistance genes in the population. Metal resistance genes in particular are highly susceptible to lateral transfer and have accumulated in multiple *Rhodanobacter* species due to a putative recombinational hot spot in the *Rhodanobacter* pan genome. This ability to rapidly acquire metal resistance genes via LGT likely explains in part the ability of *Rhodanobacter* to dominate uranium-contaminated sites.

**MATERIALS AND METHODS**

**Comparison of Rhodanobacter 16S rRNA gene sequences to the OR-IFRC 100-well-survey 16S rRNA gene data.** One hundred wells from various areas selected in the OR-IFRC research area were recently sampled for 16S rRNA gene sequence diversity (45). OR-IFRC sequences were obtained using MiSeq as previously described (58). Aligned *Rhodanobacter* 16S sequences were obtained from the Ribosomal Database (46). *Rhodanobacter* and OR-IFRC samples were compared using BLASTn (47). OR-IFRC sequences resulting in BLAST hits with >97% sequence identity were extracted and counted using *ad hoc* Perl scripts.

**Abundance profiling of reference genomes and metagenomes.** Abundance profiles for the two metagenomes, available *Rhodanobacter* genomes, and all available bacterial reference genomes based on COG categories were downloaded from the IMG website (59–62). For other *Rhodanobacter* genomes, the COG abundance information was extracted directly from the associated GenBank files using *ad hoc* Perl scripts. Odds ratios (OR) for each COG were calculated using the equation OR = (A/B)/(C/D), based on the following contingency tables: A is the number of genes in the genome/metagenome of interest assigned to COG category i, B is the number of genes in the genome/metagenome assigned to all COG categories, C is the number of genes from all sequenced bacterial genomes assigned to COG category i, and D is the number of genes from all sequenced bacterial genomes assigned to COG categories. Odds ratios and significance tests were calculated using *ad hoc* Perl scripts. If C or D equaled 0, the value was changed to 1 to allow for calculation. Each contingency table was tested for significance using the χ² test (df = 1; P ≤ 0.05), and the right-tailed Fisher exact test (P ≤ 0.05) was used to identify COG categories over abundant in the selected genome/metagenome based on all sequenced bacterial genomes. The ln odds ratio values were plotted as heat maps using R version 2.15.0 (63).

**Rhodanobacter strain isolation, sequencing, assembling, and annotation.** Groundwater samples from wells FW104 and FW510 were collected on 14 November 2012 and 26 November 2012, respectively. The samples were shipped on ice overnight. Thirty-microliter aliquots of the groundwater samples were streaked on Reasoner’s 2A (R2A) agar and tryptic soy agar (TSA) plates. Samples from FW510 were streaked on R2A and TSA plates amended with 125 μM Mn. Colonies developed within a week of incubation at room temperature. Individual colonies were picked, streaked for purity tests, and regrown in the corresponding liquid medium. Overnight liquid cultures were used to extract DNA for 16S-based identification.

In addition to the nine publicly available *Rhodanobacter* genomes (including six OR-IFRC isolates) (24, 36), seven additional OR-IFRC strains were isolated from wells FW104 (four from area 3) and FW510 (three from Y12 West), which were sequenced using Illumina MiSeq at the Institute for Environmental Genomics, University of Oklahoma. In total, 16 *Rhodanobacter* genomes (13 OR-IFRC strains and 3 reference strains) were available for comparative analysis. DNA libraries were prepared using the Nextera DNA library preparation kit (Illumina, Inc., San Diego, CA) according to the manufacturer’s protocol (October 2011 version). Briefly, 50 ng DNA (20 μl at 2.5 ng/μl) was fragmented using 5 μl of Tagment DNA enzyme with 25 μl of Tagment DNA buffer. Tagmentation reactions were performed by incubation at 55°C for 5 min, followed by purification of the tagged DNA using the Zymo Clean & Concentrator-5 kit (Zymo Research, Orange, CA). Purified DNA was
eluted from the column with 25 μl of resuspension buffer. Illumina adapter and index sequences were added to the purified tagmented DNA (20 μl) by limited-cycle PCR (5 cycles) amplification with index 1 and 2 primers in a 50-μl reaction mixture volume according to the Nextera protocol. Amplified DNA was purified using 30 μl of AMPure XP beads (Beckman Coulter, Inc., Australia). The fragment size distribution of the tagmented DNA was analyzed utilizing a 2100 Bioanalyzer with a high-sensitivity DNA assay kit (Agilent Technologies, Santa Clara, CA). DNA libraries, pooled in equal volumes, were normalized to 2 nM. Sample libraries for sequencing were prepared according to the MiSeq reagent kit protocol. Amplified DNA was quality trimmed using Btrim (Btrim64 -S -q -t <file> -o <output> -a 20 -w 4 -i 180) (64), and the resulting libraries were assembled using SOAPdenovo (65) using a range of kmer sizes (37, 55, 67, and 75 bp) (SOAPdenovo-127mer all -S <config> -K <kmer> -V <prefix> -R). For each genome, the kmer size that minimized the number of contigs and scaffolds and maximized the N50 lengths and the lengths of the contigs and scaffolds was chosen for annotation (see Table S3A in the supplemental material). Assembled sequences were uploaded to the IMG expert review pipeline for annotation (see Table S3B) (66). Four of the strains (FW104-R5, FW104-T7, FW510-T8, and FW510-R10) were assembled to draft quality, while the remaining three (FW104-R3, FW510-R10, and FW510-R12) assembled poorly.

Phylogenetic tree construction. For most of the phylogenetic trees, protein alignments were constructed in MEGA 5.2 (67) using Muscle (68). The phylogenetic trees were constructed using the maximum-likelihood method based on the JTT matrix-based model (69) using a discrete gamma distribution with invariant sites allowed to model rate differences among sites. To analyze the distribution of mera genes in the OR-IPRC metagenomes, reference Mera protein sequences were obtained for sequenced reference genomes (including all available strains of Rhodanobacter) from GenBank. Only those sequences most closely corresponding to true Mera enzymes based on previous studies were maintained (70). A reference tree was created in MEGA 5.1 (67). The evolutionary history was inferred by using the maximum-likelihood method based on the JTT matrix-based model (69). The tree with the highest log likelihood (−12,814,280) is shown. The initial tree(s) for the heuristic search were obtained automatically by applying the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with superior log likelihood value. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories [+G, parameter = 1.1818]). The rate variation model allowed for some sites to be evolutionarily invariable ([I+I], 5.0071% sites). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 199 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, less than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 144 positions in the final data set. The bootstrap consensus tree was derived from 500 replicates (71). Branches corresponding to partitions that were reproduced in less than 50% of bootstrap replicates were collapsed. Metagenome sequences corresponding to putative Mera proteins were identified by BLASTp analysis (47). Because many of the metagenome fragments did not overlap, a traditional phylogenetic tree could not be constructed. Therefore, metagenome fragments were aligned to the reference alignment using the MAFFT multisquence alignment tool (72) using the following command: mafft --retree 0 --treeout --reorder --localpair --weight 0 --averagelinkage input. The final tree was formatted for publication using ITOL version 2 (73).

Detection of LGT events. Whole-genome alignments were prepared using Mauve (74) to provide a reference visualization of LGT events. Four complementary methods were used to identify putative LGT events: phylogenetic comparison (identifying deviations from the 16S tree), Alien_Hunter (interpolating variable motifs based on compositional bias), DarkHorse (lineage probability index), and GOHTAM (local variations of genome signatures). Putative LGT events were detected using the GOHTAM software (50) with the sliding windows plus codon usage option and default options. Since the draft genomes consisted of multiple contigs, all contigs were concatenated into a single sequence with appropriate adjustment of feature coordinates prior to GOHTAM analysis. The complete (closed) R. denitrificans 2APBS1 genome was screened for the presence of any candidate laterally transferred genes via Alien_Hunter (49) and DarkHorse (48). The threshold score for the program Alien_Hunter was set as 20 instead of the default to minimize the number of false positives after observing the distribution of the scores. The resulting predictions were plotted, along with the genes involved in the denitrification pathway, heavy metal resistance, integrases, and transposases, using the program CGView (Fig. 1) (75). The program DarkHorse was built on the GenBank nr database downloaded on 30 July 2012 and was run with the default filtering threshold of 0.1 on two different "self" sequence definitions, (i) Genus: Rhodanobacter, and (ii) Species: R. sp. 2APBS1. Since the Rhodanobacter GenBank nr data are composed of sequences from the isolates from the same site that were found to display high similarity levels, the genus Rhodanobacter was also selected as "self" to avoid false negatives. A comparative distribution obtained upon using the different self terms is shown in Fig. S1 in the supplemental material.

Nucleotide sequence accession numbers. The Rhodanobacter species sequenced as part of this study were deposited in NCBI GenBank under BioSample accession numbers SAMN02929088, SAMN02929136, SAMN02929138, SAMN02929139, SAMN02929141, SAMN02929142, and SAMN02929143. The Whole Genome Shotgun projects have been deposited at DDBJ/ENA/GenBank under the accession numbers LVJP0000000-LVJV0000000. The versions described in this paper are versions LVJP0100000-LVJV0100000. All other sequences have been reported previously (1, 24, 34, 36). All sequences are also available through the JGI IMG database (http://img.jgi.doe.gov/).

SUPPLEMENTAL MATERIAL


Figure S1, TIF file, 1 MB.
Figure S2, TIF file, 0.9 MB.
Figure S3, TIF file, 1.1 MB.
Figure S4, TIF file, 1.5 MB.
Figure S5, TIF file, 2.7 MB.
Figure S6, TIF file, 1.9 MB.
Figure S7, TIF file, 1.1 MB.
Table S1, DOCX file, 0.02 MB.
Table S2, DOCX file, 0.01 MB.
Table S3, DOCX file, 0.01 MB.

ACKNOWLEDGMENTS

This material by ENIGMA-Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research, under contract number DE-AC02-05CH11231.
FUNDING INFORMATION
This work, including the efforts of Terry C, Hazen and Adam P.Ark, was funded by U.S. Department of Energy (DOE) (DE-AC02-05CH11231 and DE-FG02-07ER64398).

REFERENCES


