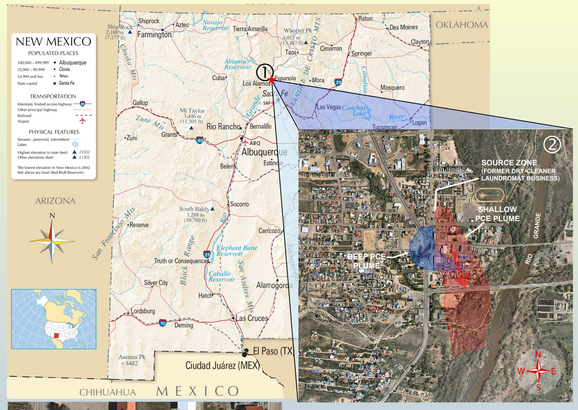
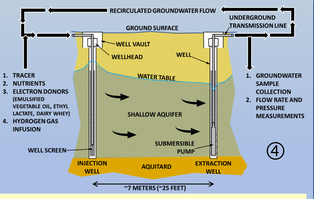
Comparison of Whole Genome Sequencing, 16S Amplicons, and qPCR for Assessment and Monitoring of an EPA Superfund Site

Peter Guerra, Principal Environmental Engineer, Lynker, Albuquerque, NM
Rebecca A. Reiss, Professor Emeritus, Biology, New Mexico Institute of Mining and Technology (NMT), Socorro, NM
Oleg Makhnin, Associate Professor, Math, NMT, Socorro, NM

Background and Project Information



The project site, remedial design, remedial action, and optimization is previously discussed in ***Guerra et al. Bioremediation of a Chlorinated Hydrocarbon Plume: A Superfund Site Field Pilot Test, Applied Sciences, 2021, 11, 10005, https://doi.org/10.3390/app112110005***



North Railroad Avenue Plume (NRAP) Superfund Site located in Espanola, New Mexico, USA ①

② Approximate groundwater plumes prior to remediation and the groundwater flow directions in the shallow aquifer (red plume and arrow) and deep aquifer (blue plume and arrow) are shown. The location of the source of contamination, a former drycleaner/laundromat business, is also shown.

③ Source and Hot-spot Areas of the shallow plume and remediation wells from which groundwater DNA samples were collected for comparison

④ Conceptual schematic of selected remedy for the Source and Hot-spot Areas consisting of groundwater recirculation with the addition of bioamendments, nutrients and hydrogen gas.

Groundwater-microbe Sample Collection

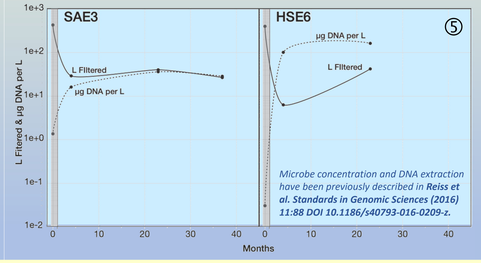
①

Well	Samples (Dates)	WGS	16S
Source Area Extraction Well 3 (SAE3)	0 months, Baseline (June, 2007)	√	√
	4 months (Nov. 2007)*	√	√
	23 months (June, 2009)	√	√
	39 months (Oct. 2010)**	√	√
Hot Spot Extraction Well 6 (HSE6)	0 months, Baseline (June, 2007)		√
	4 months after start (Nov. 2007)*	√	√
	23 months after start (June, 2009)	√	√ ④

* Listed as SAE3-5 and HSE3-5 in the JGI database
** Listed as SAE3-47 in the JGI site

②

③



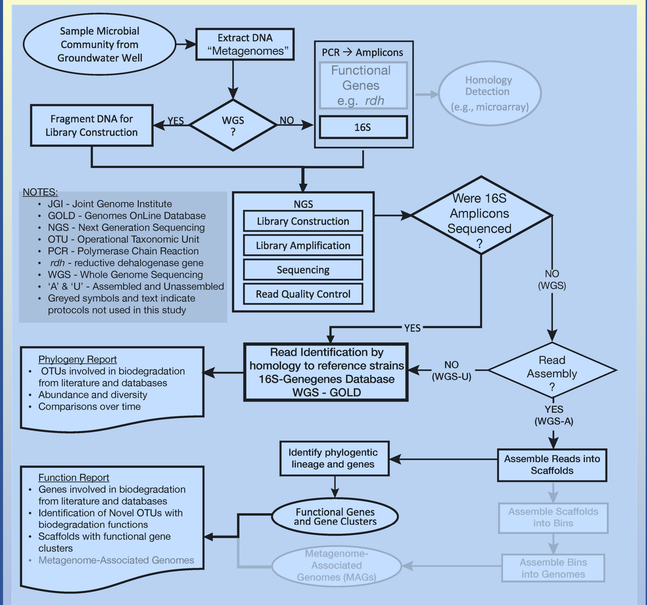
Microbe concentration and DNA extraction have been previously described in ***Reiss et al. Standards in Genomic Sciences (2016) 11:88 DOI 10.1186/s40793-016-0209-z.***

On-Site Sampling. After purging the transmission lines from the extraction wells and the extraction-well flow-control and sampling manifold①, recirculating groundwater flow was diverted② through a sterile filter capsule③ which was discharged to a graduated bucket.

④Samples were collected from two wells, SAE3 and HSE6 prior to the start of remediation and up to 39 months after treatment.

⑤ **Sampling and DNA yield Results**. For each well, the solid line is the amount of water filtered on site and the dotted line is the µg of DNA per L of water extracted. The shaded region is pre-remediation.

Environmental Metagenome Workflow

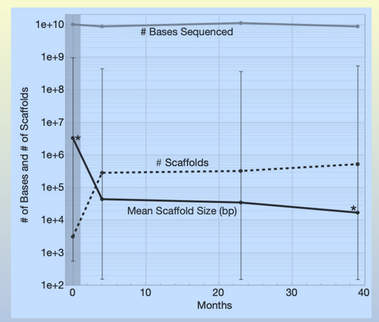


Initial microbial collection and DNA extraction are similar, except that whole WGS requires more DNA than PCR. WGS requires fragmentation of the DNA prior to the addition of adapters that include sequences from which amplification and sequencing procedures proceed. PCR starts with an amplification step that utilizes universal primers, and the resulting fragments (amplicons) that can be assigned taxa by homology comparison using hybridization techniques, such as microarrays, or undergo the same NGS process as WGS fragments. The resulting sequence data for amplicons are compared to 16S databases to identify the taxonomic group. For WGS, databases of genome reference stains are necessary to tag each read with a phylogenetic identification. Abundance, richness calculations, and visualization of results from both techniques facilitates comparison. Assembled reads creates scaffolds in which gene coding regions can be found and annotated. Steps in grey are not included in this study.

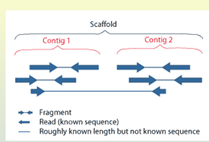
Summary of WGS-U Results

Well-months	Biomass DNA/L	# Reads Sequenced (raw data)	Filtered Raw Data	# Reads ID'd	% ID'd
Overall		644,215,230	627,783,262	18,591,015	3.0%
SAE3-00	1.35	95,204,046	92,029,566	173,596	0.2%
SAE3-04	16.05	111,898,856	109,931,044	2,511,640	2.3%
HSE6-04	100.8	130,851,632	128,284,954	2,944,944	2.3%
SAE3-23	36.08	107,826,190	105,748,376	3,347,195	3.2%
HSE6-23	162.39	116,961,818	112,008,068	5,142,674	4.6%
SAE3-39	28.39	81,472,688	79,781,254	4,470,966	5.6%

Summary of WGS-A Results

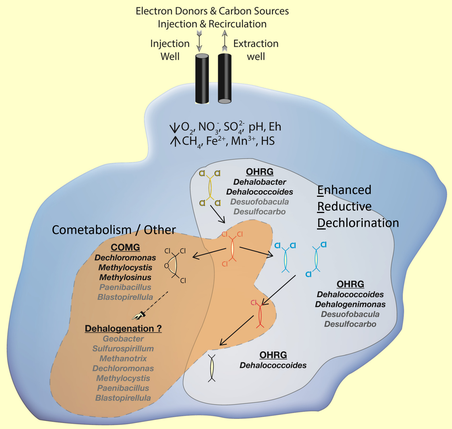


The solid grey line are the total number of bases for each time point, the black dotted line indicates the number of scaffolds at each time point and the solid black line is the mean scaffold size. The bars indicate the range of scaffold size, asterisks indicate significantly different mean scaffold size (One way ANOVA, Tukey HSD, 0.05) and the shading indicates pre-treatment.

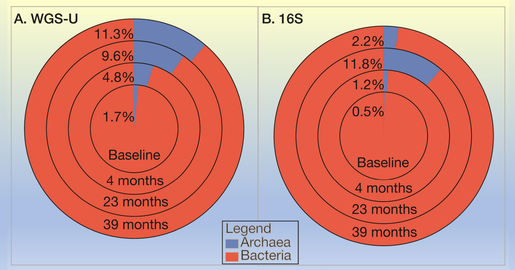


What is a Scaffold?
A scaffold is a portion of the genome sequence reconstructed from end-sequenced whole-genome shotgun clones. Scaffolds are composed of contigs and gaps. A contig is a contiguous length of genomic sequence in which the order of bases is known to a high confidence level. Gaps occur where reads from the two sequenced ends of at least one fragment overlap with other reads in two different contigs (as long as the arrangement is otherwise consistent with the contigs being adjacent). Since the lengths of the fragments are roughly known, the number of bases between contigs can be estimated.

https://mycocosm.jgi.doe.gov/help/scaffolds.jsf



Domain Level Abundance by Method

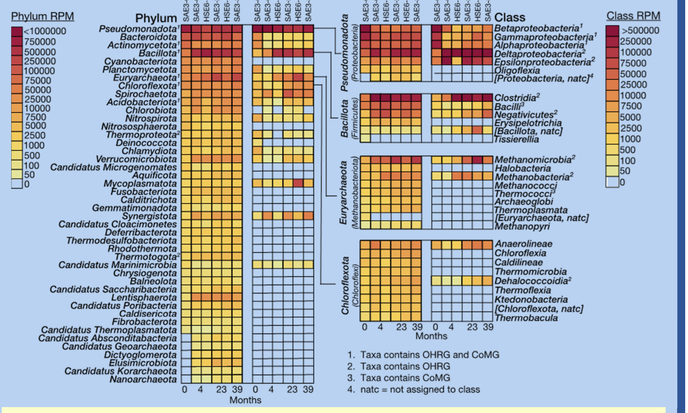


Percentages of Archaea are shown.

Taxa Richness by Method

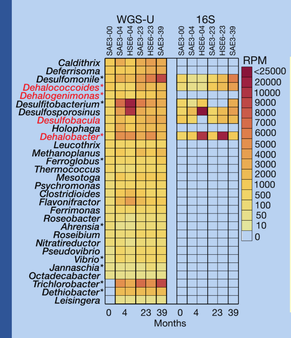
	Phylum			Class			Order			Family			Genus		
Method	16S	WU[1]	WA[2]	16S	WU	WA	16S	WU	WA	16S	WU	WA	16S	WU	WA
Sample															
Overall	18	57	50	39	115	123	73	216	257	132	459	564	237	1514	2560
SAE3-0	17	51		35	103		61	197		99	428		156	1368	
SAE3-4	17	44		31	92		55	177		85	377		134	1153	
HSE6-4	16	44		32	92		54	177		83	377		133	1154	
SAE3-23	17	44		33	92		58	177		97	377		155	1155	
HSE6-23	17	44		35	92		58	177		91	376		143	1153	
SAE3-39	16	44		33	92		55	177		105	377		178	1155	

Phyla and Selected Class-Level Results



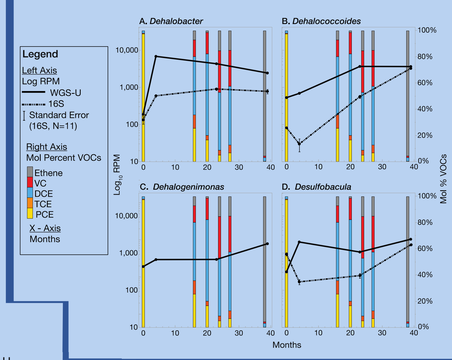
Only phyla that are present post-remediation are shown, those in brackets are not assigned to a class (natc). Data are sorted on abundance in WGS baseline (0 months). Superscripts indicate taxa known to contain literature-validated dehalogenating members

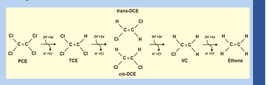
Organohalide Respiring Genera



Function-Informed Phylogenetics was used to select strains containing *rdh* genes in the GOLD database using IMG (Intregrated Microbial Genomes) and cross-referencing the WGS-U list. The abundance of these genera is based on the genomes that contain rdh genes in the GOLD database. Those in red have *rdh* genes in the NRAP dataset. Asterisks indicate literature-validated Organohalide Respiring Genera (OHRG).

Select OHRGs in Relation to PCE and Its Respiration to Ethene

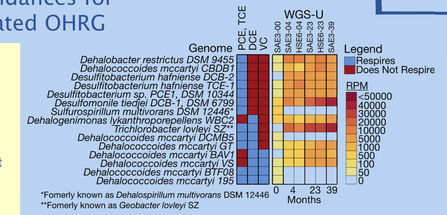




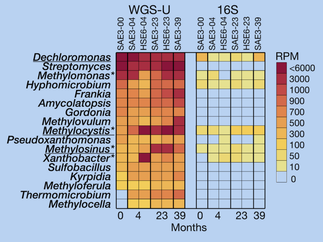
OHRGs in relationship to Dissolved PCE Concentrations and Its Biodegradation Byproducts
A. Dehalobacter (Phylum Bacillota),
B. Dehalococcoides (Chloroflexota) and C. Dehalogenimonas (Chloroflexota) are validated in the literature as OHRG and scaffold(s) with *rdh* genes are present in WGS-A data.
D. Desulfobacula is not yet known to be an OHRG, but a scaffold containing a *rdh* gene is present in WGS-A.

Strain-level Abundances for Literature-validated OHRG

These genomes have the ability to respire various chloroethenes and include *rdh* genes in the sequenced reference strains. The abundance of reads attributable to each strain at NRAP is based on unassembled WGS reads.

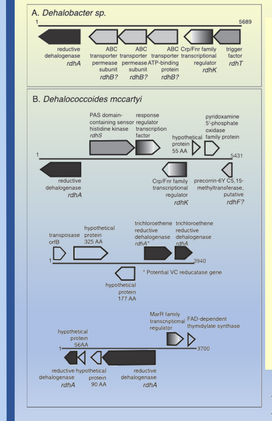


Potential Co-metabolizing Genera



These genera have genomes with either MMO or TMO genes and increase in abundance over time in WGS-U. Those underlined have scaffolds containing MO genes in the WGS-A data The asterisks indicate genera validated in the literature as being able to catalyze the oxidation of TCE.

Select NRAP *rdh* Containing Scaffolds



An assembly that included WGS for all timepoints and both wells reveals 384 scaffolds containing dehalogenase genes. 20 of the 384 were *rdh*-containing scffolds. Two species annotated with *rdh* genes, Desulfocarbo indianensis and Desulfobacula phenolica, are not yet known to be involved in chlorinated-solvent biodegradation. Multiple genes with functions that support reductive dichlorination were identified in the WGS-A scaffold analysis. The length of the scaffold are indicated by the numbering.

A. Dehalobacter scaffold with genes transcribed in the same direction, indicative of an operon.

B-D. Dehalococcoides scaffolds. Scaffold C is notable since it contains two potential VC reductase genes and a transposase gene.

NOTES:
1. Arrows indicate direction of transcription
2. Scaffold length in base pairs is shown