SUPPORTING INFORMATION for

Proteomic analysis of ethene-enriched groundwater samples from a vinyl chloride-contaminated site

Adina S. Chuang[1], Yang Oh Jin[1], Laura S. Schmidt[1], Yalan Li[2], Samuel Fogel[3], and Timothy E. Mattes[1]*

[1]Department of Civil and Environmental Engineering, University of Iowa, Iowa City, IA
[2]University of Iowa Proteomics Facility, Iowa City, IA
[3]Bioremediation Consulting, Inc. Watertown, MA

*Corresponding author email: tim-mattes@uiowa.edu

SUMMARY

Number of pages: 12

Number of figures: 3

Number of Tables: 3

Supplementary methods

Extraction of DNA and proteins from microcosms

Cells in sample pellets used for protein extraction were lysed with one of the following methods: (1) sonication in an ethanol ice bath with 10 second pulses for 3 minutes at the 45% amplitude setting using a 3-mm microtip on a Sonic Dismembrator 550 (Fisher Scientific, Tustin, CA), (2) three cycles through a cold (~4°C) French press mini-cell (124,000 kPa) (Thermo Scientific, Waltham, MA), or (3) mixing with 1 mL of 6 M guanidine and 10 mM DTT in 50 mM Tris/10 mM $CaCl_2$ [pH 7.6], 60°C. To avoid contamination betweens samples, the French Press and Sonic Dismembrator were rinsed thoroughly between extractions with ethanol, Windex (SC Johnson, Racine, WI), and deionized water. For guanidine-treated samples, a 6-fold dilution with 50 mM Tris/10 mM $CaCl_2$, pH 7.6, was performed prior to SDS-PAGE or trypsin digestion. Following lysis, the soluble fraction of protein extracts was separated by centrifugation (21,000xg for 5 minutes at 4°C).

PCR, cloning, and DNA sequencing of EtnC and EtnE genes from microcosms. PCR products were purified with the Qiagen PCR purification kit, cloned into the pDRIVE vector (Qiagen) and transformed into competent *E. coli* (New England BioLabs). Transformants were grown on LB+kanamycin (50 ng/µl) with 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (X-gal) (40 ng/µl), and Isopropyl-β-D-Thiogalactopyranoside (IPTG) (100 mM) at 37°C overnight. Single white colonies were transferred into LB+Kanamycin (50 ng/µl) broth, and incubated at 37°C overnight

with shaking (250 rpm). Recombinant vectors were purified with the QIAprep Spin Miniprep Kit (Qiagen) and sequenced with the M13 forward primer.

Proteomics analyses

SDS-PAGE. Gel-resolved samples were first denatured in Laemmli Sample Buffer (Biorad Laboratories, Hercules, CA) for 5 minutes at 95°C per manufacturer's instructions and loaded onto polyacrylamide gels (Criterion Tris-HCl Gel, 12.5%, Biorad Laboratories, Hercules, CA). Electrophoresis was conducted until clear bands of polypeptide separation were observed after Coomassie Blue staining (200 V, 55 min).

Trypsin digestion. After excision, gel slices were washed with 50% 100 mM NH_4HCO_3/50% acetonitrile (ACN) and then with 50% 10 mM NH_4HCO_3/50% ACN to remove Coomassie stain. Gel pieces were dried by centrifugal evaporation and treated with 20 mM DTT (1 hour, 56°C) followed by a 55 mM iodoacetamide solution (1 hour, 25°C, dark conditions) to prevent formation of disulfide bonds. Gel slices were then washed with two cycles of first 25 mM NH_4HCO_3 followed by 50% 25 mM NH_4HCO_3 /50% ACN and then dried to completeness in a vacuum centrifuge (Savant SpeedVac Concentrator SPD111V-115, Thermo Scientific). Sequencing-grade trypsin (1:50 w/w) (Promega Corp., Madison, WI) in 25 mM NH_4HCO_3 was added to samples for overnight digestion at 37°C. The following day, the supernatant (containing peptides) was removed for MS analysis. To improve peptide extraction efficiency,

digested gel slices were sonicated, washed with 50% H_2O/50% ACN, and all supernatants combined. All samples were concentrated by vacuum centrifugation to approximately 5 μL.

Peptide separation by SCX chromatography. As an alternative to SDS-PAGE separation, two protein extracts were digested with trypsin in solution, and the resulting peptides separated with strong cation exchange (SCX) chromatography. The soluble protein fraction was digested with trypsin overnight (1:50 w/w) at 37°C, treated with 20 mM DTT for 1 hour at 37°C, desalted and exchanged into 0.1% formic acid in HPLC-grade water using SepPak Lites C_{18} (Waters Corp., Milford, MA), and concentrated by centrifugal evaporation to approximately 500 μL. Peptides were filtered through an Ultrafree-MC centrifugal filter unit (Millipore Corp., Billerica, MA), separated using SCX on UltraMicrospin or Microspin columns (Nest Group, Inc., Southboro, MA), and eluted with a step gradient of ACN/ammonium formate buffers (*Buffer 1*: 25% ACN/75% 10 mM ammonium formate, pH 4.0; *Buffer B*: 25% ACN/75% 200 mM ammonium formate). The elution steps were 0%, 10%, 20%, 30%, 40%, 50%, 60%, 100% Buffer B in Buffer A. Peptide fractions (50-100 μL) were desalted with PepClean C-18 Spin Columns (Thermo Scientific) and concentrated by centrifugal evaporation to approximately 5-10 μL for mass spectrometry analysis.

Tandem mass spectrometry of peptides. All resulting peptides were then subjected to liquid chromatography coupled to electrospray-tandem mass spectrometry (LC-ES-MS/MS). Peptides were loaded onto a New Objective ProteoPepII column (5 μm, RP-C18 resin, New Objective, Woburn, MA) using an Eksigent nanoLC autosampler connected to a Thermo Finnigan LTQ XL

mass spectrometer. All samples were analyzed using a 75-minute reverse phase gradient from 95% H_2O/5% ACN 0.1% formic acid (FA) to 30% H_2O/70% ACN/0.1% FA. The LTQ mass spectrometer was operated as follows: nanospray voltage (2.4 kV), heated capillary temperature (200°C), full scan m/z range (400-2000). Data dependent MS/MS spectra were collected with the following parameters: 8 MS/MS spectra for the most intense ions from the full scan (minimum peak signal intensity required =3000.0 counts, isolation width=3.0) with 35% collision energy for collision-induced dissociation. Dynamic exclusion of the same abundant peptides was enabled with a repeat count of 1 and an exclusion duration of 1 minute or 3 minutes. LC columns were washed several times with the above described gradient or replaced between sample sets to avoid contamination between runs. Between analysis of samples from different wells, the MS/MS spectra of buffers flowing through a cleaned column were visually inspected to ensure that only masses one would expect from a blank run were present. This was done to further minimize the possibility of crossover contamination.

Protein sequence database development. The "biomarker database" contains the following sequences: *etnC* and *etnE* clones sequenced in the present study (Genbank Acc. GQ847806-GQ84782), complete EtnC and EtnE sequences from VC-assimilating *Nocardioides* sp. strain JS614 (Genbank Acc. AAV52084 and AAV52081), *Mycobacterium* sp. strains JS60 and JS623 (Genbank Acc. AAO48576, AAO48573, ACM61847, ACM61843); partial EtnE sequences from VC-assimilating *Mycobacterium* sp. strains JS61, JS616, JS617, JS619 (Genbank Acc. AAO48578-AAO48581), ethene-assimilating *Mycobacterium* sp. strains JS622, JS624, and JS625 (Genbank Acc. AAO48583, AAO48585, AAO4586), VC-assimilating *Pseudomonas* sp. strain AJ and

Ochrobactrum sp. strain TD (Genbank Acc. AAX52771); complete EaCoMT (XecA) sequences from propene-assimilating *Gordonia rubripertincta* strain B-276 (Genbank Acc. AAL28081) and *Xanthobacter* strain Py2 (Genbank Acc. Q56837); partial EtnC sequences from ethene-assimilating *Mycobacterium* sp. strains NBB1, NBB2, NBB3, and NBB4 (Genbank Acc. ABB70482-ABB70485); and 70 SDIMO alpha subunit sequences (Genbank Acc. ABB70412–ABB70481). To increase the database size and contents to better analyze false positive protein identifications, two human gut isolate genomes ((*Anaerofustis stercorihominis* DSM 17244 (NZ_ABIL02000006) and *Clostridium leptum* DSM 753 (NZ_ABCB00000000))), two common soil bacteria genomes (*Geobacter bemidjiensis* Bem (CP001124) and *Shewanella sediminis* HAW-EB3 (CP000821)), and common contaminants (www.thegpm.org/cRAP/index.html) were added to the biomarker database.

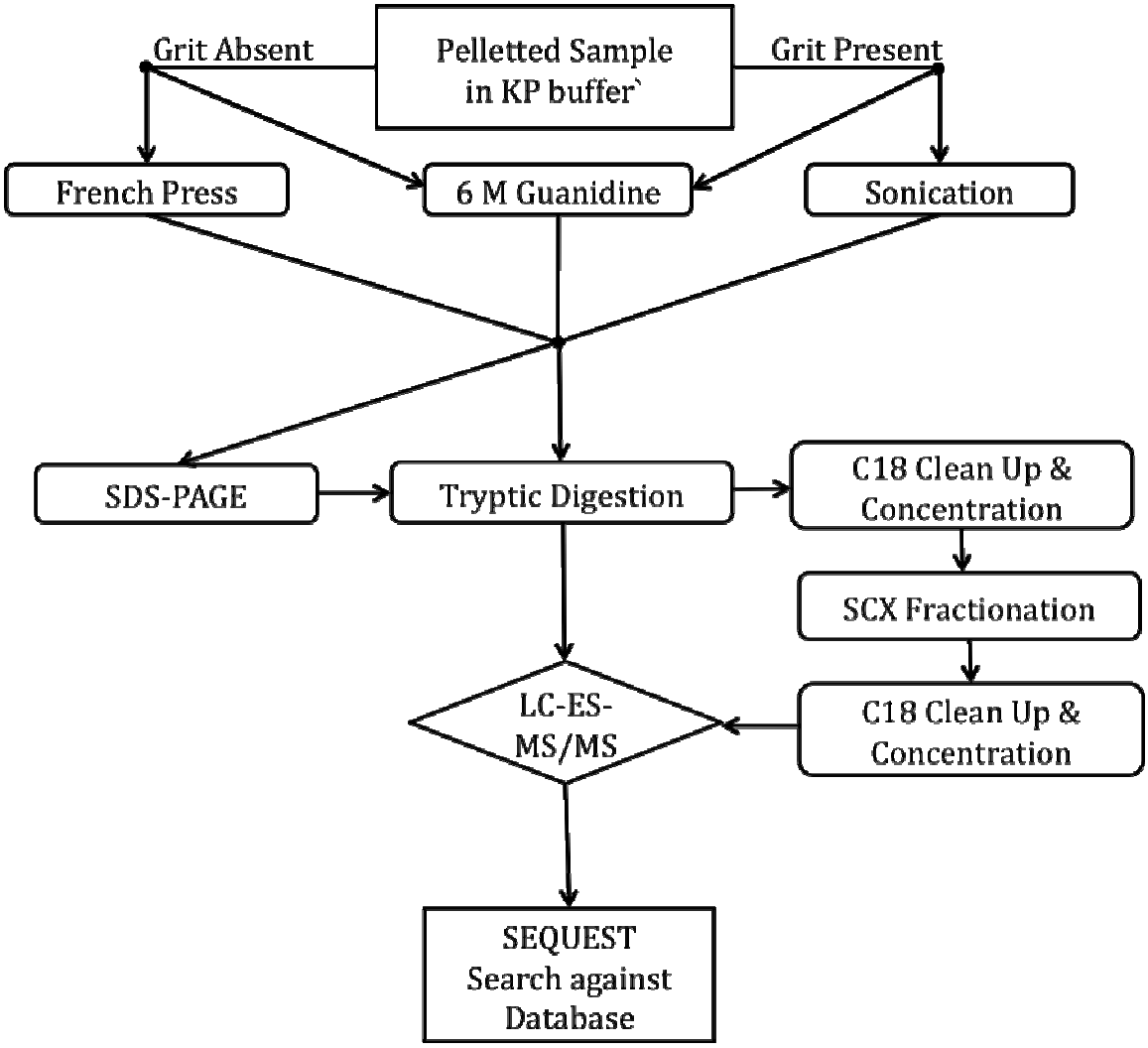

Figure S1. Workflow for protein identification in enrichment samples.

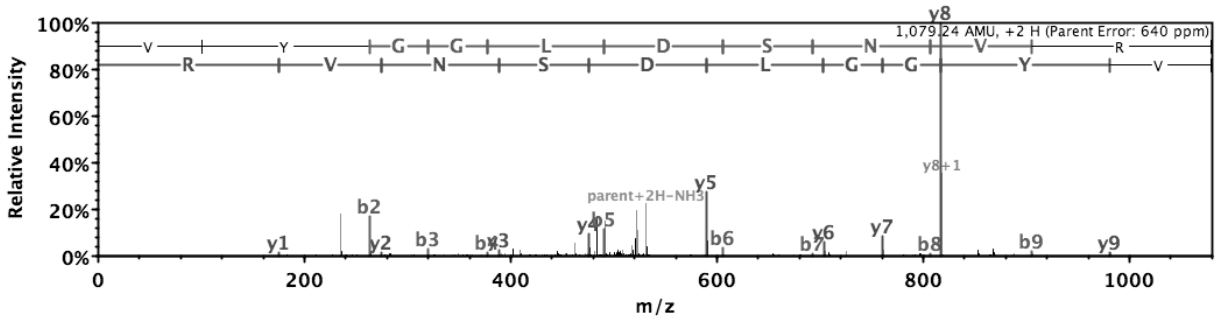

Figure S2. Representative spectra of an EtnC peptide identified from the RB-73 microcosm (Scaffoldv2_06_00, Proteome Software Inc., Portland, OR).

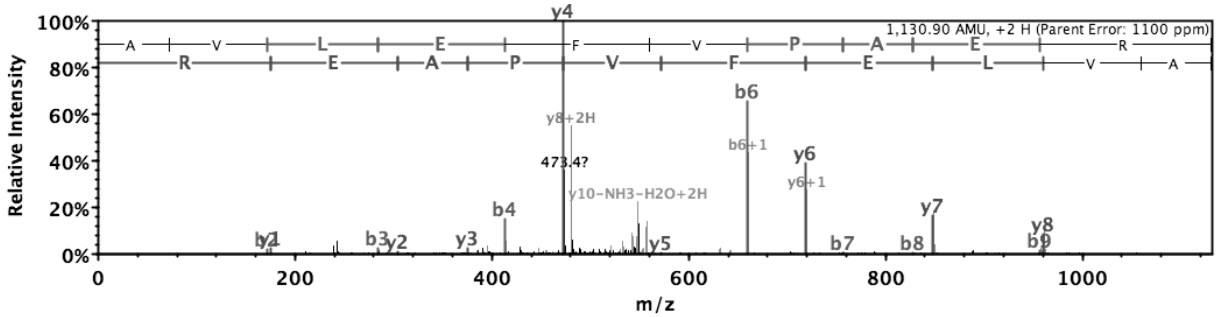

Figure S3. Representative spectra of an EtnE peptide identified from the RB-63I microcosm (Scaffoldv2_06_00, Proteome Software Inc., Portland, OR)

Sample	Peptides Identified	SEQUEST XCorr	SEQUEST deltaCn	Observed m/z	Actual Peptide Mass (AMU)	Charge
RB-46	***NANLAEPR***^{JS614}	2.70	0.23	443	884	2
RB-46	***VYGALDSNVR***^{JS614}	3.78	0.47	547	1093	2
RB-46	FMEAMK	2.13	0.15	756	755	1
RB-60	LEVVHEK	2.55	0.28	428	853	2
RB-60	GYAAMESEK	2.23	0.30	985	984	1
RB-60	MWTIDDIR	2.84	0.40	527	1051	2
RB-60	QGYAGQMLDEVR	2.81	0.34	684	1366	2
RB-60	VYGGLDSNVR	2.92	0.30	540	1078	2
RB-60	LSPFGLKPPAR	2.89	0.53	591	1181	2
RB-60	DVWEEWVVDDFVGSYMSR	2.63	0.56	1111	2221	2
RB-73	QGYAGQMLDEVR	2.67	0.51	685	1367	2
RB-73	LEVVHEK	2.57	0.33	428	853	2
RB-73	FIEGMK	2.71	0.35	428	855	2
RB-73	HTQIEVALR	2.66	0.34	526	1050	2
RB-73	MWTIDDIR	2.52	0.25	540	1078	2
RB-73	VYGGLDSNVR	2.04	0.31	1050	1049	1
RB-73	TLIGQPHLNAER	3.29	0.54	675	1348	2
RB-73	IVYGEGK	1.65	0.22	766	765	1

Table S1. EtnC Peptides identified in enrichments by searching spectra against the biomarker database. Peptides were filtered by requiring XCorr values of at least 1.5 (+1), 2.5 (+2), 3.5 (+3), and delta Cn values of greater than 0.08. Bold, italicized peptides are unique to a particular EtnC sequence (source strain indicated in superscript) when searched with Blastp against the NCBI nr database.

Sample Name	Peptides Identified	SEQUEST XCorr	SEQUEST deltaCn	Observed m/z	Actual Peptide Mass (AMU)	Charge
RB-46	ALADGAAIVR	3.35	0.47	479	957	2
RB-46	DFVAGIIDVK	3.33	0.41	540	1077	2
RB-46	***AVLEFVPAER***^{JS614}	2.74	0.34	567	1132	2
RB-46	STVTETADEVADR	3.83	0.56	698	1395	2
RB-52I	STVTETADEVADR	3.77	0.31	697	1393	2
RB-52I	LGLSTDCGLINLPR	4.64	0.53	765	1529	2
RB-58I	AFNEDFK	2.83	0.43	436	869	2
RB-58I	ALADGAAIVR	3.57	0.42	480	957	2
RB-58I	LGLSTDCGLINLPR	3.70	0.41	511	1530	3
RB-58I	DFVAGIIDVK	2.93	0.34	540	1078	2
RB-58I	***AVLEFVPAER***^{JS614}	2.61	0.40	565	1129	2
RB-58I	STVTETADEVADR	4.21	0.56	698	1394	2
RB-58I	ELTAQQA	1.54	0.38	761	760	1
RB-58I	EAFEDAVGAIVHDQEAAGLDIISDGR	6.63	0.60	900	2697	3
RB-58I	TGPDDLKPLVDNNWEK	3.38	0.38	922	1843	2
RB-60	GEFVYDAISR	3.10	0.39	579	1156	2
RB-60	***SVLEFVPAEQLALSTDCGLINLNR***^{JS623}	3.75	0.51	887	2659	3
RB-60	EAFEDAVLAIVHDQEAAGLDIISDGK	3.82	0.37	910	2726	3
RB-63I	AFNEDFK	3.22	0.49	479	956	2
RB-63I	DFVAGIIDVK	2.86	0.46	540	1078	2
RB-63I	STVTETADEVADR	1.82	0.20	607	605	1
RB-63I	ALADGAAIVR	4.55	0.63	688	1375	2
RB-63I	***AVLEFVPAER***^{JS614}	4.63	0.58	699	1395	2
RB-63I	EAFEDAVGAIVHDQEAAGLDIISDGK	4.20	0.57	765	1528	2
RB-63I	***IGSQDVLLPTTMVGNYPNPR***^{JS614}	2.19	0.25	871	870	1
RB-63I	LGLSTDCGLINLPR	5.38	0.58	891	2669	3

RB-63I	MISAGK	3.03	0.13	929	1856	2
RB-63I	TGPDDLKLLVDNNWEK	1.94	0.42	957	956	1
RB-63I	WYDGSGFATFPK	2.80	0.29	1076	1075	1
RB-63I	EAFEDAVGAIVHDQEAAGLDIISDGR	6.63	0.60	900	2697	3
RB-63I	TGPDDLKPLVDNNWEK	3.38	0.38	922	1843	2
RB-63I	PSDAPAPQAAVPIFPEVLGANIDALNYEVGR	5.64	0.70	1065	3192	3
RB-64I	ALADGAAIVR	3.35	0.47	479	957	2
RB-64I	DFVAGIIDVK	3.33	0.41	540	1077	2
RB-64I	***AVLEFVPAER***^{JS614}	2.74	0.34	567	1132	2
RB-64I	STVTETADEVADR	3.83	0.56	698	1395	2
RB-73	EFGAWVLDK	3.94	0.40	579	1156	2
RB-73	GEFVYDAISR	3.53	0.34	579	1156	2
RB-73	LHALADGAAIVR	3.35	0.36	579	1157	2
RB-73	STITETADEVADR	3.45	0.39	580	1158	2
RB-73	***VGSTDVLLPTTMVGNYPNPR***^{JS623}	2.80	0.34	580	1159	2
RB-73	AAAIFPK	4.63	0.55	1066	2131	2
RB-73	PFVAGVIDVK	4.37	0.41	1067	2132	2

Table S2. EtnE Peptides identified in enrichments by searching spectra against the biomarker database. Peptides were filtered by requiring XCorr values of at least 1.5 (+1), 2.5 (+2), 3.5 (+3), and delta Cn values of greater than 0.08. Italicized peptides are unique to only one strain when searched with Blastp and the NCBI nr database. Bold, italicized peptides are unique to a particular EtnE sequence (source strain is indicated in superscript) when searched with Blastp against the NCBI nr database.

Sample	Uniprot Accession	Protein Function Annotation	JS614 Gene Number
RB-46, RB-60	Q5U9K9	Probable alkene monooxygenase alpha subunit, EtnC	Noca_4807
RB-46, RB-52I, RB-58I, RB-63I, RB-64I	Q5U9J8	Epoxyalkane: coenzyme M transferase, EtnE	Noca_4810
RB-46	A1SC62, A1SC91	Coenzyme A transferase	Noca_4813
RB-46	Q5U9K1	Putative short-chain dehydrogenase	Noca_4814
RB47I, RB-60, RB-63I	A1SC71	Succinate-semialdehyde dehydrogenase (NAD(P)(+))	Noca_4822
RB-47I	A1SC76	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Noca_4827
RB-47I, RB-60	A1SC82	Succinate-semialdehyde dehydrogenase (NAD(P)(+))	Noca_4833

Table S3. Proteins identified in enrichments that are associated with the VC/ethene biodegradation pathway in *Nocardioides* sp. strain JS614. Spectra were searched against the proteins encoded by the JS614 genome. A successful protein identification required XCorr values of at least 1.5 (+1), 2.5 (+2), 3.5 (+3), dCn values of at least 0.08; and at least two unique peptides. Protein IDs are UniProt accession numbers (www.uniprot.org). Gene numbers were generated during machine annotation of the JS614 genome sequence (genome.ornl.gov/microbial/noca).