Comparative Metagenomics of Freshwater Microbial Communities

R-2923

C. Hemme[1,2], Y. Deng[1,2], Q. Tu[1,2], M. W. Fields[3], T. J. Gentry[4], L. Wu[1,2], S. G. Tringe[5], D. B. Watson[6], Z. He[1,2], T. C. Hazen[7], J. M. Tiedje[8], E. Rubin[5], J. Zhou[1,2]

[1]Univ. of Oklahoma, Norman, OK, [2]Inst. for Environmental Genomics, Norman, OK, [3]Montana State Univ., Bozeman, MT, [4]Texas A&M Univ., College Station, TX, [5]Joint Genome Inst. Production Genomics Facility, Walnut Creek, CA, [6]Oak Ridge Natl. Lab., Oak Ridge, TN, [7]Lawrence Berkeley Natl. Lab., Berkeley, CA, [8]Michigan State Univ., East Lansing, MI.

ENIGMA

Ecosystems and Networks Integrated with Genes and Molecular Assemblies

http://vimss.lbl.gov/

Introduction

Previous analyses of a microbial metagenome from uranium and nitric-acid contaminated groundwater (FW106) showed significant environmental effects resulting from the rapid introduction of multiple contaminants. Effects include a massive loss of species and strain biodiversity, accumulation of toxin-resistant genes in the metagenome and lateral transfer of toxin resistance genes between community members. To better understand these results in an ecological context, a second metagenome from a pristine groundwater system located along the same geological strike was sequenced and analyzed (FW301).

It is hypothesized that FW301 approximates the ancestral FW106 community based on phylogenetic profiles and common geological parameters; however, even if is not the case, the datasets still permit comparisons between healthy and stressed groundwater ecosystems. Complex carbohydrate metabolism has been almost entirely lost in the stressed ecosystem. In contrast, the pristine system encodes a wide diversity of complex carbohydrate metabolism systems, suggesting that carbon turnover is very rapid and less leaky in the healthy groundwater system. FW301 encodes many (~160+) carbon monoxide dehydrogenase genes while FW106 encodes none. This result suggests that the community is frequently exposed to oxygen from aerated rainwater percolating into the subsurface, with a resulting high rate of carbon metabolism and CO production. When oxygen levels fall, the CO then serves as a major carbon source for the community. FW301 appears to be capable of CO2 fixation via the reductive carboxylase (reverse TCA) cycle and possibly acetogenesis, activities; these activities are lacking in the heterotrophic FW106 system which relies exclusively on respiration of nitrate and/or oxygen for energy production. FW301 encodes a complete set of B12 biosynthesis pathway at high abundance suggesting the use of sodium gradients for energy production in the healthy groundwater community.

Overall comparative analysis suggests that the introduction of contaminants is accompanied by a decrease in biodiversity, loss of nutrient cycling, less metabolic diversity, increased respiration and dominance by *r*-strategists. These results are consistent with trends predicted for stressed ecosystems.

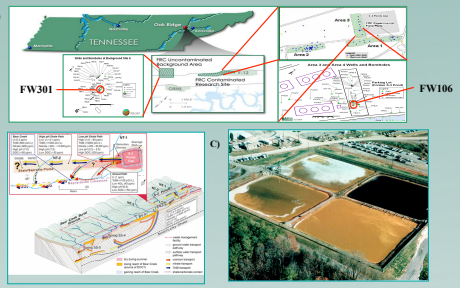


Fig. 1. A) Geographic location of sampling wells FW106 and FW301. B) Model of contaminant flow from the former S-3 waste disposal ponds. C) Photo of former S-3 waste disposal ponds.

FRC Groundwater Geochemistry

	FW300	FW106
pH	~7	~3.7
Nitrate (mg/L)	1.5	2331
Sulfate (mg/L)	6.3	1997
Uranium (mg/L)	>0.0001	51
Technetium-99 (pCi/L)	-	3700
1,2-Dichloroethene (µg/mL)	5	1153
Tetrachloroethene (µg/mL)	5	810
1-Butanol (µg/mL)	-	475
Acetone (µg/mL)	10	823
Benzoic Acid (µg/mL)	-	1400
Sodium (mg/L)	2.96	826
Chloride (mg/L)	1.125	465
Magnesium (mg/L)	2.58	45.7

FRC Metagenome Statistics

	FW106	FW301	Minnesota Farm Soil	Lake Washington Sediment (combined)
Total Bases	9554544	106573620	152406385	211470570
DNA Coding # Bases	8076611	84170988	97186775	173308284
DNA G+C # Bases	6011119 (63.20%)	60905616 (57.16%)	85978723 (57.65%)	123326011 (58.34%)
Scaffolds	5698	126845	135347	241455
Total # Genes	12420	170521	185274	323777
16S rRNA	3	51	23	64
Genes w/ Function Prediction	8689 (69.96%)	86161 (50.53%)	80680 (43.55%)	185893 (57.41%)
Genes Assigned to Enzymes	3692 (13.62%)	53901 (31.61%)	21907 (14.46%)	16773 (5.18%)
Genes Assigned to KEGG Pathways	3423 (11.46%)	44007 (25.81%)	18310 (9.88%)	14000 (4.32%)
Genes in COGs	7061 (84.10%)	85262 (50.00%)	85014 (45.89%)	174344 (53.85%)
Genes in Pfam	7410 (59.66%)	79392 (46.56%)	13913 (7.51%)	166747 (51.50%)

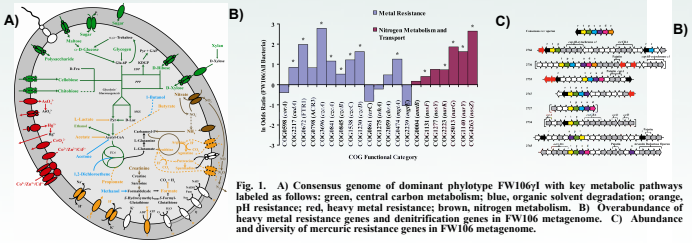


Fig. 1. A) Consensus genome of dominant phylotype FW106y1 with key metabolic pathways labeled as follows: green, central carbon metabolism; blue, organic solvent degradation; orange, pH resistance; red, heavy metal resistance; brown, nitrogen metabolism. B) Overabundance of heavy metal resistance genes and denitrification genes in FW106 metagenome. C) Abundance and diversity of mercuric resistance genes in FW106 metagenome.

Full results for FW106 metagenome characterization: Hemme, et al., 2010 *ISME J.*, 4:660-672

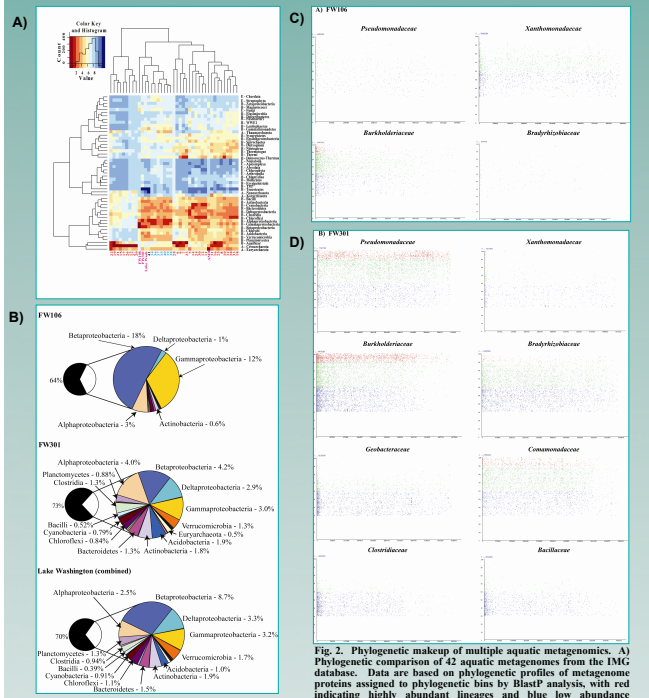
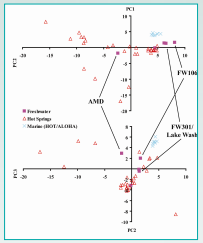


Fig. 2. Phylogenetic makeup of multiple aquatic metagenomics. A) Phylogenetic comparison of 42 aquatic metagenomes from the IMG database. Data are based on phylogenetic profiles of metagenome proteins assigned to phylogenetic bins by BlastP analysis, with red indicating highly abundant lineages and blue low abundance lineages. B) Pie chart of same data as displayed in A showing similarities between FW106, FW301 and Lake Washington fresh water sediments (combined). Protein recruitment plots for C) FW106 and D) FW301.



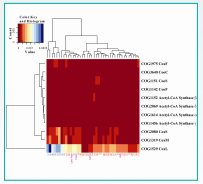


Fig. 3. A) PCA analysis of 42 aquatic metagenomes based on COG profiles demonstrating similarities in COG profiles of FRC and Lake Washington samples. B) Abundance of carbon monoxide dehydrogenase genes in aquatic metagenomes.

A)

	COG	Description
FW106	COG1291	Cytochrome *bd* quinol oxidase subunit 2
	COG2009	Succinate dehydrogenase/fumarate reductase cytochrome *b* subunit
	COG2010	Cytochrome *c* monoheme/diheme variants
	COG2193	Bacterioferritin (cytochrome b_{1})
	COG2863	Cytochrome c_{551}
	COG2993	cbb_3 cytochrome oxidase, cytochrome subunit
	COG3038	Cytochrome b_{561}
	COG3125	Heme/Copper-type cytochrome/quinol oxidase subunit 4
	COG3245	Cytochrome c_4
	COG3658	Cytochrome *b*
	COG3909	Cytochrome c_{550}
	COG4654	Cytochrome c_{554}/c_{552}
FW301	COG5274	Cytochrome *b* involved in lipid metabolism
MFS	COG3005	Nitrate/TMAO reductase membrane-bound tetraheme cytochrome *c* subunit
	COG3303	Formate-dependent nitrite reductase periplasmic cytochrome c_{552} subunit
	COG4736	cbb_3 cytochrome oxidase subunit 3
AMD	COG1290	Cytochrome *b* subunit of *bc* complex
	COG1969	NiFe hydrogenase I cytochrome *b* subunit

B)

Function ID	Name	AMD	LW	MF S	FW30 1	FW10 6
COG1319	Aerobic-type CODH, CoxM/CutM homologs	2	84	60	49	0
COG1529	Aerobic-type CODH, CoxL/CutL homologs	8	298	199	250	4
COG2080	Aerobic-type CODH, CoxN/CutN subunit	3	139	60	71	1
COG1052	CODH/Acetyl-CoA synthase, α-subunit	2	5	0	0	0
COG1614	CODH/Acetyl-CoA synthase, β-subunit	0	12	1	0	0
COG2069	CODH/Acetyl-CoA synthase, δ-subunit	0	4	1	1	0
COG1456	CODH/Acetyl-CoA synthase, γ-subunit	0	18	3	0	0
COG2640	CODH maturation factor (CODH nickel-insertion accessory protein CooC)	1	10	1	1	0
COG1051	6Fe-4S primase cluster-containing protein (CODH catalytic subunit CooS)	1	19	0	0	0
COG1162	Fe-S-cluster-containing hydrogenase components 2 (CODH Fe-S subunit CooF)	0	13	2	3	0
COG1975	Xanthine and CODH Maturation factor, XdhC/CoxF Family	3	70	34	34	3

Table 1. A) Abundance of cytochromes in FW106, FW301, soil (MFS) and acid mine drainage (AMD). Cytochrome c_{552} is highly abundant in FW106 and may contribute to heavy metal resistance. B) Abundance of carbon monoxide dehydrogenase genes in various metagenomes.

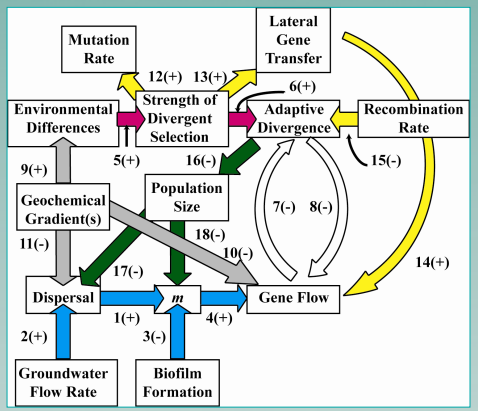


Fig 4. Model of gene flow and adaptive divergence in contaminated groundwater systems. Adapted from Räsänen and Hendry, 2008. *m* refers to the proportional migrant population, +/- signs indicate positive and negative effects, respectively. Colored arrows are as follows: cyan, naturally-occurring factors in groundwater ecosystems affecting gene flow; magenta, naturally-occurring factors affecting adaptive divergence; white, feedback loop between gene flow and adaptive divergence; grey, factors affecting gene flow and adaptive divergence based upon externally-imposed geochemical gradients; yellow, naturally-occurring and stress-related genetic and genomic factors affecting adaptive divergence and gene flow; green, demographic factors affecting adaptive divergence and gene flow. + and − indicate positive or negative effects, respectively.

- **FRC groundwater communities are dominated by proterobacterial species**
- **Introduction of strong geochemical gradients drastically reduce species, strain and metabolic diversity, leading to reduced nutrient turnover, increased nutrient loss, increased community respiration and increased lateral transfer of geochemical resistance genes**
- **A strong geochemical gradient may act as a lens do reduce biodiversity, limit gene flow and increase the odds of emergence of adaptive strains.**

ACKNOWLEDGEMENTS

ENIGMA is a Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.