Spatio-Temporal Feedbacks Between Soil Contamination and Biotic Responses in LNAPL Zones Undergoing Natural Source Zone Depletion

Matthew Morgan | Research Scientist – CSIRO Contaminants and Biotechnology Program

GQ19, Liège, Belgium. 12 September, 2019

Geoffrey Puzon; Melanie Bruckberger; Matthew Taylor; Jian-Wei Lui; Shoko Okada; Tom Walsh; Trevor Bastow; John Raynor; Greg Davis



www.csiro.au

Legacy contamination site

- Largely weathered petroleum hydrocarbon contaminants
 - > 50 years old in many cases
- LNAPL subsurface contaminants consisting of Jet fuel/Kerosene, diesel and crude oil
- Previous work on active remediation methods (i.e. air sparging, biosparging and multiphase extraction)
- Currently investigating Natural Source Zone Depletion (NSZD) processes and details of the microbial function and activity in NSZD





Aim & Approach

- Identify direct and indirect evidence of natural attenuation of soil contaminants by microbial communities through analysis of functional diversity, biological activity, metabolic pathway analysis, and detection of degradation products
- Who is there?
 - Community composition 16S metabarcoding, shotgun metagenomics
- What can they do?
 - What genes and pathways are present DNA Shotgun Metagenomics
- What are they doing?
 - Realised functional profile: Proteomics, Metabolomics
- Integrate multi-omic signatures in order to model interactions and predict community metabolic outcomes
 - Transfer generic outcomes to other sites, with appropriate consideration of site and environmental variation.



Soil core sampling

- From each core 4 depth sections were chosen
- 1. Top section (unsaturated minimal contamination)
- 2. Water Table / Capillary fringe
- 3. Mid section/ high TPH (saturated and contaminated)
- 4. Bottom (saturated and largely uncontaminated)
- Cores were sterilely cut into 0.05m sections, sealed in tins and placed on ice for transport
- In the lab, cores were subsamples for chemical analysis (TPH) and molecular analysis (metagenomics, proteomics and metabolomics)
- All samples processed in triplicate (DNA)





TPH profile within cores

GC FID Profiles



Contaminant Fingerprints





Community comparisons

- Inferred taxonomic composition and community structure compared across cores and depths
- High similarity of biological replicates
- Significant effects of contaminant and depth on community composition
- Strong separation: TPH vs no TPH
- Top samples relatively similar to each other and to Background
 - Jet fuel top section samples distinct from others





Inferred Functional Potential





- Similar structure of functional potential and taxonomy across cores and depths
- Similarity of Top section and Background samples



Proteomics

- Protein extracted from core samples
 - Significant challenge
 - Jian-Wei Lui
- Gel shows good extraction of proteins from all contaminated samples and background
- Proteins were excised from the gel and run on ThermoFisher Orbitrap Fusion





Proteomics analysis

- Created database derived from metagenome predicted ORFs
 - 6,575,581 predicted open reading frames
 - 6,560 total proteins identified
 - 1,899 annotated to a known function
- Many unknown function or provenance
 - Can still analyse distribution of "anonymous" proteins to discover candidate biomarkers for further characterisation



Realised functions

- Normalised protein expression across cores and depth (no biological replication)
- Significant effects of contaminant and depth on community functional composition
- Top section samples more similar to contaminant samples than to Background
 - Functional potential similar to Background
 - Realised functions more similar to other contaminated soils





Proteomics Results

- High level results largely consistent with inferred functional potential in contaminated samples
 - Confirmed methanogens present and active in contaminated soils
 - Methanogenesis proteins highly expressed in contaminated samples, including coenzymeF420
 - Proteins from degradation pathways (e.g. naphthalene and toluene) are active, and more highly expressed in contaminated soils than expected from metagenomic abundance
- Different patterns in distribution of potential and realised functions in Top section samples
 - Chloroplast and photosystem I & II proteins and transporters highly expressed in Background samples (especially Top section).
 - Not detected in Top section contaminated samples
- Active taxa and pathways performing highly expressed functions may not be highly abundant in metagenomics data



Discriminant metabolites



 23 of 145 identified metabolites are significant in discriminating the samples including degradation products of naphthalene (salicylic acid) and crude/diesel (4-hydroxybenzoate)

Analyte ID put.	d.value	stdev
Phenylacetic acid-TMS	486.29	4493500000
Thymine-2TMS	298.67	1650600000
Uracil-2TMS	293.27	4770100000
Triethanolamine-3TMS	172.43	4.18E+11
Methylsuccinic acid-2TMS	172.1	645280000
Lysine-4TMS	160.61	46876000
3-Methylglutaric acid-2TMS	157.28	153530000
Phenylalanine-2TMS	117.03	8 8.89E+10
2-Aminoethanol-3TMS	107.88	2.05E+12
Maltose-meto-8TMS1	102.95	9.79E+12
Nicotinic acid-TMS	98.417	2920500000
Adenine-2TMS	97.405	404130000
5'-Methylthioadenosine-2TMS	90.607	23706000
4-Aminobutyric acid-3TMS	89.263	4.10E+10
Tyramine-3TMS	83.113	578450000
Adenosine-4TMS	82.036	6 4.39E+10
Putrescine-4TMS	78.557	1911500000
Palmitic acid-TMS	74.468	3.78E+12
Inositol-6TMS2	72.58	5.00E+11
Sucrose-8TMS	70.464	1.57E+13
3-Hydroxyisobutyric acid-2TMS	63.502	1.24E+13
Salicylic acid 2TMS RT15.605	60.329	625230000
4-Hydroxybenzoic acid-2TMS	55.506	2475100000



Summary

- Developed new protocols for multi-omic extractions from samples with high contaminant levels from disparate sources
- Preliminary analyses are establishing relationships within and between omics data layers
- Direct and indirect evidence of active biological attenuation pathways
- Different patterns in potential and realised functions demonstrate value of realised function assessment for integrating data layers into metabolic models



Ongoing Work

- Longitudinal sampling to support temporal analyses
- Characterising MAGs
 - Genome-scale metabolic modelling for species and communities
 - Agent-based dynamic models versus steady-state
- Integrated multi-omic pathway analysis
- Inform modelling to understand how community dynamics relate abundance and activity
 - Potential for modulation and manipulation



Thank you

Matthew Morgan

CSIRO Land and Water Research Scientist Phone: +61 2 6246 4172 Email: matthew.morgan@csiro.au

LAND AND WATER www.csiro.au



Inferred taxonomic composition

- Taxonomy explored at multiple resolutions
 - Class is relatively conservative
- Dominant bacterial taxa relatively similar across cores and depths
- Similarity of biological replicates
- Strong separation of TPH vs no TPH largely related to Archaeal methanogens present in contaminated samples







Pathway analysis

- KEGG pathways inferred for 23 significantly different compounds
- Seven compounds identified in 'microbial metabolism in diverse environments' pathway ID





Phenylalanine metabolism – good correlation with two metabolites for sample 116





4-hydroxybenzoate

- Breakdown product of p-cresol
- P-cresol is found in diesel and crude oil

https://www.chemicalbook.com/Ch emicalProductProperty JP CB5453 502.htm https://pubs.acs.org/doi/10.1021/a c00130a021





Discriminant Taxa



CSIRC

Discriminant Functional Potential



Methanogenic taxa and proteins identified

- Example from one • protein fraction of Core D6 (Diesel contaminant)
- *Methanosaeta* proteins • are highly abundant in diesel samples
- But not abundant in metagenomics data





Species



