An Examination of the Issues Associated with Analyzing Complex Microbial Communities

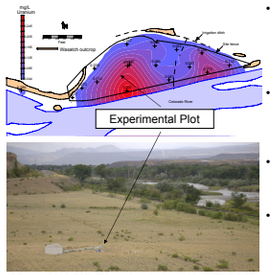
comparing an in-silico approach to an applied approach

Carrie D. Nicora, Angela D. Norbeck, Philip E. Long, Mary S. Lipton, Heather M. Mottaz, and Richard D. Smith
Pacific Northwest National Laboratory, Richland, WA 99352

Overview

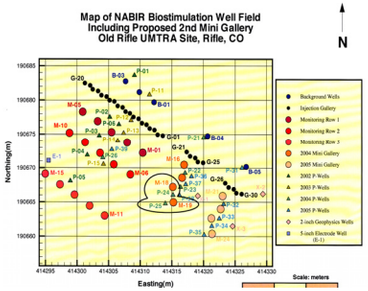
- Global proteomics offers a unique opportunity to study the relationships between microorganisms by examining changes in species composition and protein expression on a community level. This examination process has, however, proven to be difficult due to many levels of complexity.
- A high throughput proteomic method for peptide identification detecting species-specific protein expression (using MS/MS) was explored in two ways. Initially through an *in-silico* study, and second through experimental measurements.
- The method application was applied to a current bioremediation site. Samples were taken from a contaminated Colorado River sediment community on a former uranium ore-processing site, before and after bio-stimulation.
- A significant problem in assessing the approach was the efficiency of removing particle associated sediment bacteria. Emphasis on the use of enzymes and physical disruption to liberate cells for analysis was probed.
- While results from our *in-silico* study illustrated the feasibility of detecting an organism against a background of other organisms in a microbial community, the applied approach demonstrated many complexities which hampered the ability to discern individual peptide identifiers.
- Issues included: soil extraction, analysis and informatics problems.

Introduction



Map of NABIR Bio-stimulation Well Field (1)

- Certain dissimilatory metal-reducing microorganisms are capable of reductively precipitating radioactive metals in contaminated subsurface environments. The precipitation of heavy metals decreases the risk to organisms by lowering its ability to leach into waterways. (1)
- In order to study the microbial processes that make bioremediation possible, the microbes must be removed from the sediment, the proteins extracted and digested into peptides and then analyzed on a mass spectrometer.
- The purpose of the *in-silico* study was to assess the utility of peptide identification methods for detecting a single species within a community of species.
- For the *in-silico* study, proteins predicted from annotated genomes that correspond to 25 microorganisms were compiled, virtually digested, and analyzed to determine whether a particular peptide could be used as a unique identifier in the presence of the other peptides. (2)

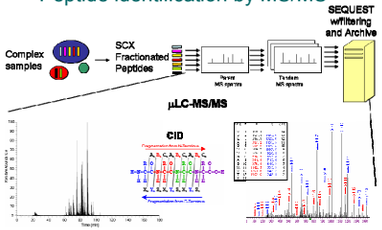


Map of Sample Locations (1)

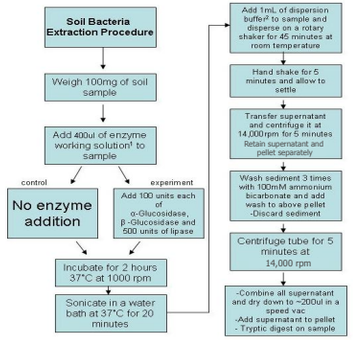
- For the application study, samples were collected from an UMTRA (Uranium Mill Tailing Remedial Action) site in Rifle Colorado, before and after biostimulation.
- Seven soil bacteria samples were extracted from the soil with the use of light detergents and enzymes.
- To test the efficiency of the enzyme extraction method, the samples were digested in quadruplicate and each digestion analyzed on an LTQ in triplicate.

Methods

Peptide identification by MS/MS

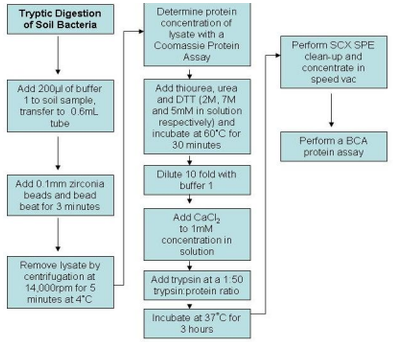


Soil Bacteria Extraction Procedure (3)



[1] Enzyme working solution – 500mM Tris/500mM Sodium Phosphate buffer, 2mM $MgCl_2$ (pH 7.3)
[2] Dispersion buffer – 7.5mM Sodium pyrophosphate, 5mM EDTA, 0.1% Chaps (pH 8.2)

Tryptic Digest Procedure



Buffer 1 – 100mM Ammonium Bicarbonate, pH 8.4

Results

- In theory, species-specific peptides can be detected within a community and differentiation of protein expression is possible on a community level, as demonstrated with an artificial 25-organism study. (2)

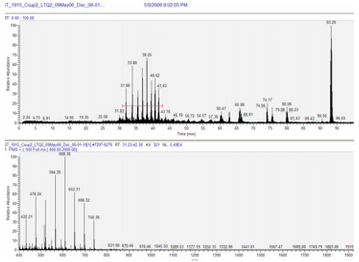
	Peptides			Proteins with a Minimum of 2 Unique peptides		
	6 ppm & 5% NET	6 ppm & 1% NET	1 ppm & 1% NET	6 ppm & 5% NET	6 ppm & 1% NET	1 ppm & 1% NET
Organism	Percent Unique	Percent Unique	Percent Unique	Percent Unique	Percent Unique	Percent Unique
B burgdorferi	2%	9%	30%	20%	78%	97%
C burnetii	2%	9%	33%	17%	72%	96%
C muridarumi	2%	9%	33%	22%	80%	98%
c pneumoniae	2%	10%	34%	19%	78%	97%
C tepidum	2%	9%	33%	16%	71%	96%
D desulfuricans	2%	10%	35%	21%	79%	99%
D Radiodurans	2%	10%	34%	19%	76%	99%
E Coli	1%	8%	27%	17%	73%	98%
G metallireducens	2%	9%	32%	19%	68%	93%
G sulfurreducens	2%	9%	31%	22%	77%	98%
M jannaschii	2%	8%	30%	21%	79%	99%
N meningitidis	2%	9%	33%	18%	73%	97%
P gingivalis	2%	9%	33%	22%	78%	97%
R palustris	2%	9%	33%	18%	76%	98%
S agalactiae	2%	9%	31%	18%	74%	97%
S aureus	2%	10%	33%	19%	70%	97%
S mitis nctc	1%	8%	30%	18%	69%	95%
S oneidensis	2%	9%	35%	18%	70%	96%
S pneumoniae	1%	6%	21%	18%	71%	98%
S typhimurium	1%	8%	28%	18%	75%	98%
T denticola	2%	9%	33%	23%	80%	98%
T maritima	2%	9%	31%	23%	82%	99%
T pallidum	2%	9%	33%	21%	79%	98%
W pipientis	2%	9%	32%	18%	73%	97%
Y pestis	2%	9%	33%	18%	75%	97%

Percentage of uniquely-identifiable peptides within an artificial complex community. With a metagenome we can characterize a community (2)

- The actual application of this approach, however, demonstrated some significant problems that likely resulted from issues introduced during all stages: sample preparation/soil bacteria extraction, basic instrument analysis issues and informatics.

Soil Bacteria Extraction Challenges:

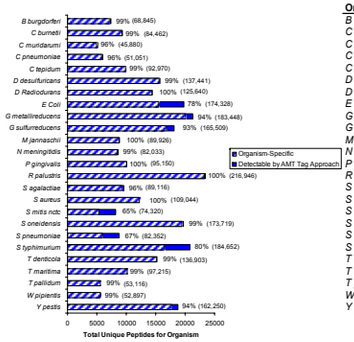
- An enzymatic bacterial extraction technique by Bockelmann, *et al.* (3) was explored to limit the use of heavy detergents that have historically been used in soil microbe extraction which are problematic for mass spectrometric analysis. Bacteria living in sediments develop biofilms, which are primarily composed of extracellular polymeric substances that adhere bacteria to soil surfaces. The enzymatic extraction method did not release the bacteria from the biofilm, resulting in low protein yield.
- The buffers suggested in the study (500mM sodium phosphate buffer for a working solution and 7.5 mM sodium pyrophosphate for a dispersion buffer), caused sodium adducts in the spectra (peaks at 44 Da apart).



Spectra of a soil sample showing the disodium peaks differing by 44 Da.

Informatics Challenges:

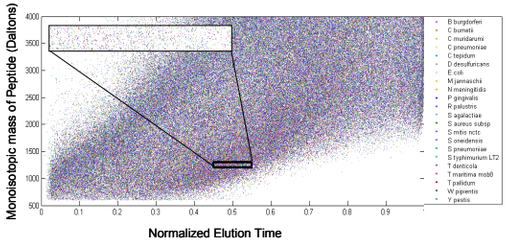
- Complexity - Multiple copies of the same protein introduce redundancy of peptide to protein identifications, and multiple organism protein files are generally appended together in database searching, in effect lowering the scores associated with best fit by commonly used search engines (delCN in SEQUEST) because of the increased numbers of possible matches. In terms of protein counts, a multiple-organism community can easily achieve a proteome of 40,000 proteins – the number currently associated with many eukaryotic systems.
- Lack of database - It is unknown exactly which microbes are present in the sample, therefore, without an annotated genomic database, it's hard to make a good identification. A match is only as good as the database.



Contributions of each organism to the total pool of unique peptides. The solid bars are the total number of peptides from each organism contributing to the pool of high confidence unique peptides from all organisms. The hashed bars represent the subset of peptide sequences that are unique to one organism. (2)

Organism	Proteins	In-silico Peptides
B burgdorferi	1637	68,845
C burnetii	2051	84,462
C muridarumi	916	45,880
C pneumoniae	1116	51,051
C tepidum	2241	92,970
D desulfuricans	3007	137,441
D radiodurans	3117	125,640
E coli	4289	174,328
G metallireducens	4767	183,448
G sulfurreducens	3449	165,509
M jannaschii	1780	89,926
N meningitidis	2047	82,033
P gingivalis	1916	95,150
R palustris	4833	216,946
S agalactiae	2115	89,116
S aureus	2676	109,044
S mitis nctc	2148	74,320
S oneidensis	4798	173,719
S pneumoniae	2098	82,352
S typhimurium	4550	184,652
T denticola	2756	136,903
T maritima	1842	97,215
T pallidum	1030	53,116
W pipientis	1176	52,897
Y pestis	4286	162,250
Totals	66,641	2,829,213

Protein and peptide counts for an artificial, complex microbial community of 25 organisms, in alphabetical order by organism. Fully tryptic peptides with at most one missed cleavage site were considered. (2)



Global representation of all unique peptides from the complex community. Microbial communities can be as complex as eukaryotic systems. (2)

Conclusions

- While the *in-silico* study revealed the potential for identifying unique peptides to a specific organism from a community of organisms, with a community of 25 sequenced organisms there is high likelihood of identifying peptides.
- With a metagenome, complexity can be overcome and identification is likely. Work is in progress to develop such data sets.
- However, there are still a number of obstacles to be overcome:
  1. Soil bacteria extraction – need to identify an efficient extraction method that does not rely on heavy detergents, or sodium buffers. Current enzymatic extraction methods do not sufficiently release bacteria from biofilms.
  2. Sample processing and LC-MS analysis – working with low quantities of complex mixtures, dynamic range issues, and obtaining high quality spectra despite contamination issues.
  3. Informatics – Complexity can surpass that of the most complex eukaryotic systems, lack of a metagenome for each sample, protein homogeneity (between organism) issues, need for high throughput "de novo" spctrum analysis tools..

Acknowledgements

This research was supported by the Natural and Accelerated Bioremediation Research Program (NABiR), Office of Biological and Environmental Research, Experimental portions of this research were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

References

1. Long, P.E. et al., Acceleration of Field-Scale Bioreduction of U(VI) in a Shallow Alluvial Aquifer: Temporal and Spatial Evolution of Biogeochemistry, Nabir PI Meeting, April 26, 2005, Virginia. <http://www.pnl.gov/nabir-umtra/index.stm>
2. Norbeck, A.D.; Callister, S.J.; Monroe, M.E.; Navdeep, J.; Elias, D.A.; Lipton, M.S.; Smith, R.D. 2006. Proteomic Approaches to Bacterial Differentiation. Journal of Microbial Methods. (in press).
3. Bockelmann, U.; Szewzyk, U.,; Grohmann, E. 2003. A new enzymatic method for the detachment of particle associated soil bacteria. Journal of Microbial Methods. 55:201-211.
4. Finneran, K.T.; Anderson, R.T.; Nevin, K.P.; Lovley, D.R. 2002. Potential for bioremediation of uranium-contaminated aquifers with microbial U(VI) reduction. Journal of Soil Contamination. 11:339-357.

Contact Information

Carrie D. Nicora
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: carrie.nicora@pnl.gov
Phone: 509.376.0118