Modular denitrification in microbial communities using sequence analysis

Alburuj R. Rahman[1], James Shapleigh[1], Armanda Roco[1], Joseph Yavitt[2]

Department of Microbiology[1], Department of Natural Resources[2], Cornell University

Introduction

Denitrification is one way bacteria carry out nitrate respiration (5) and can be described as the reduction of nitrate to nitrogen gas, in four reduction steps each requiring a specific nitrogen oxide reductase (Fig. 1).

Earlier studies have shown that most denitrifiers possess all four reductases needed to carry out each step of denitrification. New research suggests the denitrifier community consists of many incomplete denitrifiers, organisms that carry out only certain steps of denitrification (5).

Denitrification is common in soil ecosystems. One of which is Hubbard Brook (1,2), an experimental forest that contains six watersheds used for carrying out ecosystem studies.

PURPOSE:

Metagenomics will be used to examine the denitrifying community and assess the presence of both complete and partial denitrifiers. Sequenced DNA will be analyzed in order to determine nitrogen oxide reductase distribution and test for the presence and importance of modularity, the idea that incomplete denitrifiers carry out denitrification.

NO_3^- (nitrate) $\xrightarrow{\text{(Nap)}}$ NO_2^- (nitrite) $\xrightarrow{\text{(Nir)}}$ NO (nitric oxide) $\xrightarrow{\text{(Nor)}}$ N_2O (nitrous oxide) $\xrightarrow{\text{(Nos)}}$ N_2 (nitrogen gas)

Fig. 1: Denitrification shown as four steps of reduction, from nitrate to nitrogen gas.

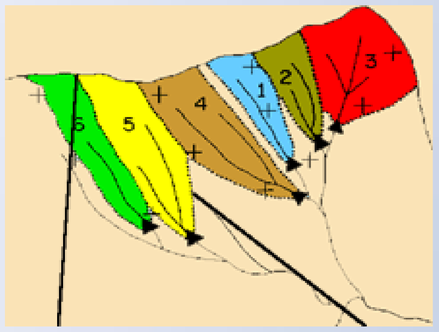

Fig. 2a: A representation of Hubbard Brooke Experimental Forest's watersheds.

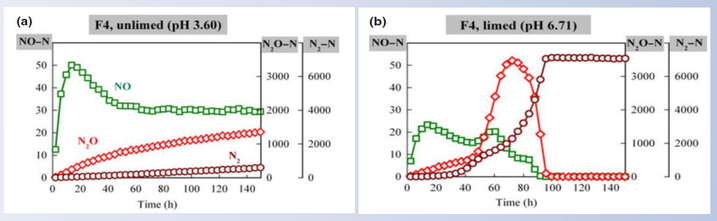

Fig. 2b: Effect of low pH on denitrification (6)

Methods

Soil was collected from two watersheds, one supplemented with Ca to raise pH. At each watershed, samples were taken from three elevations and at three depths per sample site.

DNA sequenced from samples through:
- PowerSoil DNA Isolation kits (4)
- DNA sequenced using Hi-Seq Illumina platform
- DNA processed into 180 million 100 nucleotide reads

Using the CLC, a high-throughput sequence analysis tool, data was:
- trimmed into 18-30 x 10^6 reads;
  assessed quality of reads

These trimmed paired reads were then:
- sent to a metagenomics automated analysis platform: MG-RAST (3), and re-blasted using PLAN (BLAST navigator)
- separately analyzed with blastx and blastn using custom databases

Developed programs to:
- parse hits from blastx for organismal classification
- match sequences hits from queries to original trimmed reads
- re-blast sequences to find associated organisms and assoc. biological function
- find the frequency for which a sequence from one sample would occur (have similar identity) to those in the other 17 sites (Fig. 6)

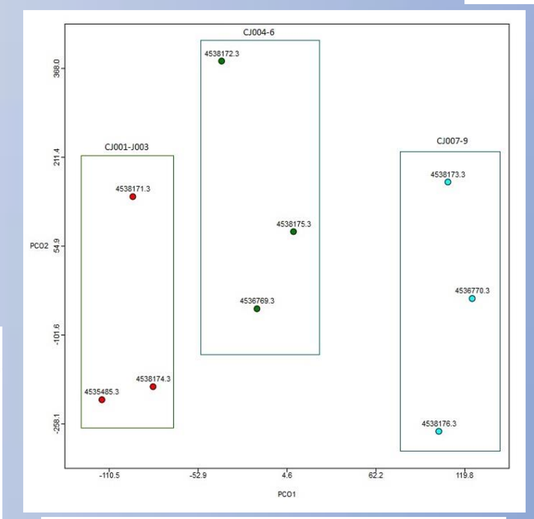

Fig. 3: A principal coordinate analysis (PCA), which examines variance in data using eigenvalues, that shows grouping of samples based on elevation.

Results

As an initial approach to establish differences between communities, a principle coordinate analysis, which examines variance in data using eigenvalues, was used for find correlation among samples. It showed clustering by elevation (Fig. 3) and by depth (Fig. 4).

For denitrification analysis, there were more hits to Nor than to Nir. Also, the most common organisms with hits to Nor, were not found to have hits to Nir.

Through blastx searches using custom Nap and Nir databases, different organismal classification was found than that from MG-RAST.

Lack of sufficient similarity in nucleic acid comparisons and inconsistencies with translated alignments and paired ends led to use of the translated sequence results exclusively.

Analysis revealed that there is an abundance of α-proteobacteria containing nirK, and actinobacteria and various proteobacteria for Nor.

Taxonomic classifications assigned to paired-end reads of NirK were more phylogenetically similar to one another than were those of Nor.

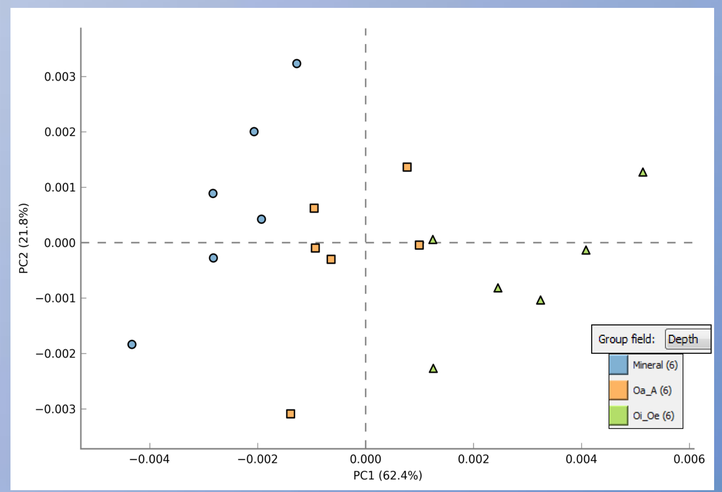

Fig. 4: PCA of samples separated by depth.

hit seq	CJ001	CJ002	CJ003	CJ004	CJ005	CJ006	CJ007	CJ008	CJ009	CJ010	CJ011	CJ012	CJ013	CJ014	CJ015	CJ016	CJ017	CJ018	TOTAL
TOTAL	828	317	148	228	372	99	211	267	81	491	198	204	493	258	117	315	254	123	5004
seq-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
seq-2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
seq-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
seq-4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
seq-5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
seq-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
seq-7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
seq-8	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
seq-9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
seq-10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
seq-11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
seq-12	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
seq-13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
seq-14	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
seq-15	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
seq-16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
seq-17	4	2	2	1	8	1	2	5	2	8	3	4	6	4	2	1	1	0	56
seq-18	6	3	1	5	3	3	5	3	1	3	2	4	4	2	2	3	7	2	59
seq-19	4	3	1	1	10	3	3	6	0	5	3	3	5	7	0	6	0	0	60
seq-20	6	9	3	2	3	2	2	1	4	7	4	5	2	1	1	3	5	1	61
seq-21	3	6	1	4	5	3	3	4	3	5	0	2	6	1	1	5	7	2	61
seq-22	5	0	2	0	5	3	0	7	1	9	5	1	5	5	2	9	4	0	63
seq-23	8	5	2	5	7	0	2	3	0	4	5	3	5	3	2	5	3	2	64
seq-24	5	5	2	4	7	1	4	1	1	8	2	6	5	4	3	4	1	1	64
seq-25	2	5	2	3	4	1	5	3	0	7	2	3	5	3	1	8	8	3	65
seq-26	6	3	1	3	8	1	3	1	1	10	4	3	11	4	2	1	2	2	66
seq-27	5	4	2	4	10	0	2	2	0	11	4	3	9	4	3	4	0	4	71
seq-28	6	4	4	4	7	2	5	4	2	6	4	4	5	4	3	3	1	4	72
seq-29	5	4	2	1	12	3	6	7	0	7	3	4	7	8	0	6	1	0	76
seq-30	3	6	3	4	5	2	5	4	1	12	2	2	7	0	1	9	9	3	78

Fig. 5: Heat-map of CJ001 sequences and frequency of its hits to the sequences of the other samples.

Discussion

A different number of hits or abundance of organisms associated with particular reductases and different organismal classifications were found to be associated with those hits (5). This goes to reinforce the notion of modularity.

The alignments will be examined through the use of semi-quantitative tests to measure differences in levels of conservation between Nir and Nor to ultimately determine which classifications can be trusted for further analysis.

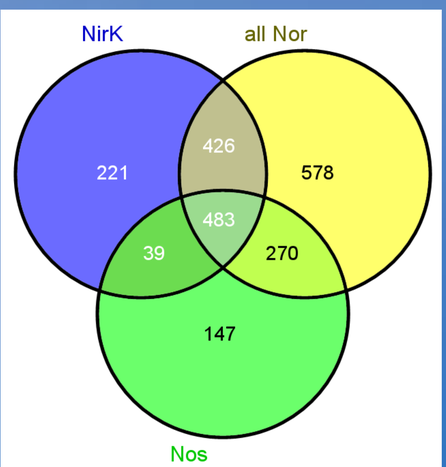

Fig. 6: nirK, nor, and nos sequenced genome distribution of the 18 samples collected.

References

1. Dittman, J.A., C.T. Driscoll, P. M. Groffman and T. J. Fahey. 2007. Dynamics of nitrogen and dissolved organic carbon at the Hubbard Brook Experimental Forest. Ecology 88:1153-1166.
2. Judd, K.E., G.E. Likens and P.M. Groffman. 2007. High nitrate retention during winter in soils of the Hubbard Brook Experimental Forest. Ecosystems 10:217-225.
3. Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., ... Edwards, R. A. (January 01, 2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. Bmc Bioinformatics, *9*.
4. Qu, Z., Wang, J., Almoy, T., and Bakken, L. Excessive use of nitrogen in Chinese agriculture results in high N2O/(N2O+N2) product ratio of denitrification, primarily due to acidification of the soils. Global Change Biology (2014) 20, 1685–1698
5. Roco, Armanda. Personal Communication. Nov. 2013.
6. Shapleigh, JP. 2012. The denitrifying prokaryotesIn E. Rosenberg, E. Stackebrandt, F. Thompson, E. DeLong, and S. Lory (eds.), The Prokaryotes. Springer, New York, NY.