

67. Detecting and quantifying genes in soil short-read metagenomes: implications for the Nitrogen cycle.

Luis H Orellana (lhorellana@gatech.edu),^{1*} Luis Rodriguez-R,² Joanne Chee-Sanford,³ Robert Sanford,⁴ Frank Löffler,^{5,6,7} and Konstantinos Konstantinidis^{1,2}

¹School of Civil and Environmental Engineering, ² and School of Biology, Georgia Institute of Technology, Atlanta, Georgia; ³US Department of Agriculture—Agricultural Research Service, Urbana, Illinois; ⁴Department of Geology, University of Illinois, Urbana, Illinois; ⁵Department of Microbiology, University of Tennessee, Knoxville, Tennessee; ⁶Bioscience Division, Oak Ridge National Laboratory, Oak Ridge; and Department of Civil and ⁷Environmental Engineering, University of Tennessee, Knoxville, Tennessee.

Project Goals:

The goals of this project are to fill existing knowledge gaps in our understanding of N- flux and associated C-turnover in soils and sediments. Novel information about the diversity, distribution, abundance and expression of genes contributing to N- transformation is required to link desirable (i.e., N-retention) and undesirable (i.e., N-loss such as that associated with N₂O emissions) activities with measurable microbial parameters. Linking molecular- and organismal-level information with environmental factors that control N- and C-turnover can predict the impact of land management practices on greenhouse gas (N₂O, CO₂) emissions. Such integrated approaches generate novel information at multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we developed and applied a bioinformatic strategy that calculates the best thresholds for detecting gene fragments related to specific genes in short-read metagenomes and use it to quantify the abundance of N-cycle genes in important soil ecosystems for bioenergy crop production in Midwest US across the year.

Abstract:

Metagenomics studies have shed light onto the natural diversity and abundance for the microbial genetic potential participating in key environmental processes. However, accurate thresholds that can discriminate between true and false positive results from commonly used similarity search methods are rarely discussed nor evaluated. To overcome these difficulties, we developed a methodology aimed to identify position- specific most-discriminant thresholds from manually curated reference datasets. Our method applies Receiver-Operator Curve (ROC)-analysis to metagenomic datasets generated in silico to simulate current short-read sequencing technologies. From the evaluation of true and false positive matches from similarity searches using simulated datasets, models with the best position-specific most-discriminating bitscore threshold for nitrogen cycle genes were constructed and evaluated. The combination of a similarity search algorithms and our strategy, showed an improved false discovery rate (FDR; from ~4 to 14 times) when compared to traditional arbitrary fixed e-value strategies. This strategy also showed to have better sensitivity (average increase of ~24%) compare to hidden markov model searches, although at the expense of more computational resources. In order to apply our tool to real metagenomic samples, we analyzed short-read metagenomes obtained from two agricultural sites with contrasting soil textures (sandy versus silty-loam) during four seasons in 2012 at two depths: surface (0-5cm) and deep (20-30 cm). Most nitrogen cycling genes (e.g., nosZ, amoA and nirK among others) varied in abundance over the course of the year. For instance, a remarkably high abundance of metagenomic reads related to atypical nosZ (reduction of N₂O to N₂) sequences were observed over the year, accounting for approximately 90% of the total nosZ reads found in both soil layers. Approximately 12% of the nosZ reads were taxonomically assigned to the Anaeromyxobacter genus, indicating their potential relevance for N₂O reduction. In addition, six amoA (ammonia oxidation) genes, each encoded by distinct archaeal and bacterial populations, became abundant in the deep sandy samples when seasonal nitrogen

fertilization was applied. This study provides a bioinformatic tool for reliable detection of target short gene fragments in metagenomes and advances our understanding of the abundance and diversity of the nitrogen cycle genetic potential found in different ecosystems. Our publicly available pipeline “ROcker” is fully automated, freely available, and can be used to investigate any other genes or processes of interest (www.enve-omics.gatech.edu).

References:

1. Robert A. Sanford, Darlene D. Wagner, Qingzhong Wu, Joanne C. Chee-Sanford, Sara H. Thomas, Claribel Cruz-García, Gina Rodríguez, Arturo Massol-Deyá, Kishore K. Krishnani, Kirsti M. Ritalahti, Silke Nissen, Konstantinos T. Konstantinidis, and Frank E. Löffler. 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc. Natl. Acad. Sci. U. S. A.* 109: 19709–14.
2. Orellana, L. H., L. M. Rodriguez-R, S. Higgins, J. C. Chee-Sanford, R. A. Sanford, K. M. Ritalahti, F. E. Löffler, and K. T. Konstantinidis. 2014. Detecting nitrous oxide reductase (*nosZ*) genes in soil metagenomes: method development and implications for the nitrogen cycle. *mBio* 5: e01193-14.

Funding Statement: This research was supported by the US Department of Energy, Office of Biological and Environmental Research, Genomic Science Program, Award DE-SC0006662.