

Integrating genomic and flux data to develop predictive models for managing methane emissions

Clifton P. Bueno de Mesquita¹ and Susannah G. Tringe^{1,2}

¹ Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, USA

² Environmental Genomics and Systems Biology Division, Lawrence Berkeley National

Focal Area:

The focal area of this white paper is the importance of integrating high potential datasets including soil genomic data, eddy covariance flux data, and remotely sensed flux data across spatial and temporal scales.

Science or Technological Challenge:

It remains a challenge to develop a mechanistic and predictive model of methane fluxes across space and time that accurately predicts how fluxes respond to environmental changes and that could be used to develop and assess emission management strategies. Such a model must incorporate microbial metabolic and ecological processes occurring at local scales that ultimately scale up to landscape-level methane fluxes.

Rationale:

Understanding microbial community taxonomic and functional composition has greatly increased our understanding of landscape-scale methane emission patterns across environmental gradients, yet predicting fluxes and how they respond to environmental change remains a major challenge [1, 2]. Across the salinity gradient in the San Francisco Estuary, we used metagenomic sequencing to elucidate the involvement of multiple microbial functional guilds and decomposition processes that drove methane emissions that were highest in oligohaline wetlands and but otherwise declined with increasing salinity [3]. A combination of metagenomic and metabolomic data also revealed halophilic methanogens contributing to the increased methane emissions observed in unrestored hypersaline salterns, a potential role for methane production by methylphosphonate-scavenging bacteria, and altered microbial community composition associated with lowered emissions after hydrologic restoration [4]. However, in a recent synthesis of methane flux and microbial data from coastal wetlands from four different sites across a wide geographic range, there were few consistencies in methane/salinity relationships and the variables driving them. Similar paired flux and microbial data from a greater number of sites would enable us to directly assess which environmental and microbial variables drive

discrepancies among observed fluxes and environmental characteristics. This, in turn, could help predict the impact of changes in ecosystem management, restoration, or other interventions..

Narrative:

A greater degree of integration between genomic and other -omic data with methane flux data is needed at expanded spatial and temporal scales. A great deal of relevant data exist or are being generated, including land- and satellite-based methane monitoring data (e.g. Ameriflux and MethaneSAT), and metagenomic and metatranscriptomic data from soils and sediments (e.g. the Integrated Microbial Genomes and Microbiomes database [5] and the National Microbiome Data Collaborative), but have not been exploited to identify microbial-methane linkages. We propose leveraging these datasets, along with metadata repositories such as the Genomes OnLine Database [6] and relevant ontologies to identify the environments, organisms, and metabolic pathways driving global methane emissions. One challenge in synthesizing these data is a lack of consistent metadata and paired microbial/methane measurements. We propose more soil sampling and microbial DNA sequencing efforts to be paired with already established methane monitoring efforts such as eddy covariance flux towers, especially in areas where such data are lacking, and the data integrated into appropriate repositories. Once enough data are generated from many sites, an analysis that synthesizes the data across sites and tests hypotheses about environmental/methane/microbial relationships would also benefit from machine learning techniques. Such techniques may include supervised machine learning models such as random forests, gradient boosting, support vector machines, ridge regression, and neural networks, or unsupervised methods [7]. These techniques will help reveal patterns in highly complex datasets comprising thousands of microbial taxa. Ultimately these data could help develop a model of the methane cycle that explicitly includes microbial processes, similar to what has been done previously with soil carbon, arid ecosystems, and other climate models [8–11]. The model could then be used to help understand which interventions, out of a variety of different options [12], would lead to the greatest reduction in methane emissions.

References

1. He S, Malfatti SA, McFarland JW, Anderson FE, Pati A, Huntemann M, et al. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. *mBio* 2015; **6**: e00066-15.
2. Hartman WH, Ye R, Horwath WR, Tringe SG. A genomic perspective on stoichiometric regulation of soil carbon cycling. *ISME J* 2017; **11**: 2652–2665.
3. Hartman WH, Bueno de Mesquita CP, Theroux SM, Morgan-Lang C, Baldocchi DD, Tringe SG. Multiple microbial guilds mediate soil methane cycling along a wetland salinity gradient. *ISME J* In review.
4. Zhou J, Theroux SM, Bueno de Mesquita CP, Hartman WH, Tringe SG. Microbial drivers of methane emissions from unrestored industrial salt ponds. *ISME J* 2021; **16**: 284–295.
5. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res* 2021; **49**: D751–D763.
6. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Sundaramurthi JC, Lee J, et al. Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Res* 2021; **49**: D723–D733.
7. Ghannam RB, Techtman SM. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Comput Struct Biotechnol J* 2021; **19**: 1092–1107.
8. Collins SL, Sinsabaugh RL, Crenshaw C, Green L, Porras-Alfaro A, Stursova M, et al. Pulse dynamics and microbial processes in aridland ecosystems. *J Ecol* 2008; **96**: 413–420.
9. Wieder WR, Bonan GB, Allison SD. Global soil carbon projections are improved by modelling microbial processes. *Nat Clim Change* 2013; **3**: 909–912.
10. Todd-Brown KEO, Hopkins FM, Kivlin SN, Talbot JM, Allison SD. A framework for

representing microbial decomposition in coupled climate models. *Biogeochemistry* 2012; **109**: 19–33.

11. Singh BK, Bardgett RD, Smith P, Reay DS. Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat Rev Microbiol* 2010; **8**: 779–790.
12. Valach AC, Kasak K, Hemes KS, Szutu D, Verfaillie J, Baldocchi DD. Carbon flux trajectories and site conditions from restored impounded marshes in the Sacramento-San Joaquin Delta. *Wetland Carbon and Environmental Management, Geophysical Monograph* 267, First Edition. 2021. John Wiley & Sons, Inc.