



Editorial: New Insights into Microbial Ecology through Subtle Nucleotide Variation

A. Murat Eren^{1,2*}, Mitchell L. Sogin² and Loïs Maignien^{2,3}

¹ Department of Medicine, The University of Chicago, Chicago, IL, USA, ² Marine Biological Laboratory, Josephine Bay Paul Center, Woods Hole, MA, USA, ³ Laboratory of Microbiology of Extreme Environments, UMR 6197, Institut Européen de la Mer, Université de Bretagne Occidentale, Plouzane, France

Keywords: oligotyping, minimum entropy decomposition, 16S rRNA gene, microbial ecology, high resolution

The Editorial on the Research Topic

New Insights into Microbial Ecology through Subtle Nucleotide Variation

Characterizing the community structure of naturally occurring microbes through marker gene amplicons has gained widespread acceptance for profiling microbial populations. The 16S ribosomal RNA (rRNA) gene provides a suitable target for most studies since (1) it meets the criteria for robust markers of evolution, e.g., both conserved and rapidly evolving regions that do not undergo horizontal gene transfer, (2) microbial ecologists have identified widely adopted primers and protocols for generating amplicons for sequencing, (3) analyses of both cultivars and environmental DNA have generated well-curated databases for taxonomic profiling, and (4) bioinformaticians and computational biologists have published comprehensive software tools for interpreting the data and generating publication-ready figures. Since the initial descriptions of high-throughput sequencing of 16S rRNA gene amplicons to survey microbial diversity, we have witnessed an explosion of association-based inferences of interactions between microbes and their environment.

Despite these advances, the field of microbial ecology faces numerous technical challenges. Sampling and storage strategies, DNA extraction protocols, limitations of the so called “universal” PCR primers, random sequencing errors, and the identification of ecologically relevant units can bias interpretations of observations based on 16S rRNA gene data. Although microbiologists comprehend most of these challenges, the need for handling large number of sequences, and to partition these complex data into appropriate proxies for environmental genomes caught almost everyone off-guard.

De novo clustering of short reads into operational taxonomic units (OTUs) based on “pairwise sequence similarities” quickly became the primary way to partition sequencing data into ecological units as this approach significantly out-performed analyses that relied strictly upon taxonomy. On the other hand, as random sequencing errors can dramatically increase the number of mismatches between two aligned reads, the underlying principle of most *de novo* clustering algorithms that rely on the edit distance was prone to inflating the diversity estimations. The use of 97% sequence similarity threshold emerged as a *de facto* standard, and has successfully reduced the impact of erroneous OTUs on diversity estimations. However, the computational convenience this arbitrary threshold offers has been at the expense of accurate ecological inference, as 3% OTUs are often phylogenetically mixed, and inconsistent (Koeppel and Wu, 2013; Eren et al., 2014; Nguyen et al., 2016).

Oligotyping (Eren et al., 2013) proposes an alternative way to decompose marker gene amplicons. It first considers the entire sequencing data to identify variable nucleotide positions,

OPEN ACCESS

Edited by:

Matthias Hess,
University of California, Davis, USA

Reviewed by:

George Tsiamis,
University of Patras, Greece

*Correspondence:

A. Murat Eren
meren@uchicago.edu

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 30 May 2016

Accepted: 09 August 2016

Published: 24 August 2016

Citation:

Eren AM, Sogin ML and Maignien L
(2016) Editorial: New Insights into
Microbial Ecology through Subtle
Nucleotide Variation.
Front. Microbiol. 7:1318.
doi: 10.3389/fmicb.2016.01318

and then utilizes only those positions that show significant variation to partition reads into oligotypes. The identification of variable nucleotide positions in the oligotyping workflow relies on Shannon entropy (Shannon, 1948), which is a measure of information uncertainty (Jost, 2006). The association between the measured entropy and the diversity of nucleotides at a given nucleotide position in a dataset of sequences allows the identification of nucleotide positions that likely carry phylogenetically important signal. The departure from pairwise sequence alignments, and the use of entropy-based decomposition strategy, makes it possible to resolve closely related but distinct taxa that differ by as little as one nucleotide at the sequenced region.

Our research topic contains original research and method papers that employs oligotyping of microbial community data to investigate ecological questions in divergent environments including the human oral cavity (Mark Welch et al.), mammalian guts (Menke et al.), deep-sea sediments (Buttigieg and Ramette), as well as freshwater (Newton and McLellan), sewage (Fisher et al.), marine (Delmont et al.), and soil (Turlapati et al.) ecosystems. In a study that cuts across multiple environments, Schmidt et al. uses oligotyping to investigate the *Vibrio* ecology in environmental, as well as host- and substrate-associated habitats. Ramette and Buttigieg implements an R package for entropy-based decomposition procedures, and their software library contains additional approaches, such as the “broken stick model” procedure to identify low-abundance oligotypes that could be generated by chance alone, and a “one-pass entropy profiling” approach to efficiently identify those OTUs whose decomposition into oligotypes would most likely explain concealed diversity (Ramette and Buttigieg). Finally, Utter et al. reconcile the individuality, stability, and variability of the oral microbial communities in the context of “spatial structure” of microbes in dental plaque by combining high-resolution depiction of microbial community data with high-resolution imaging of multi-taxa microbial consortia in the human oral cavity (Mark Welch et al., 2016).

Most articles in this collection demonstrate the importance of high-resolution analyses, and provide further evidence that reveals the need to open the “black box” of OTUs in microbial ecology. Doing so not only allows finer representation of the microbial diversity in a wide range of ecosystems, but also

improves the ecological signal for downstream analyses that aim to infer correlations (McLellan and Eren, 2014; Reveillaud et al., 2014; Eren et al., 2015; Kleindienst et al., 2015).

While oligotyping demonstrates the efficacy of an entropy-based concept to partition closely related taxa, the algorithm minimum entropy decomposition suggests that the use of information theory can be generalized to analyze entire sets of marker gene data (Eren et al., 2014; Ramette and Buttigieg). The ideal result of a properly partitioned marker gene dataset will have the minimum number of units that contains minimum entropy (i.e., none of the nucleotide positions in final units will have entropy that exceeds the expected error rate of the sequencing device), which in fact can be achieved through multiple ways. Indeed, the search for algorithms that can provide single-nucleotide resolution without relying on arbitrary percent similarity thresholds is not limited to entropy-based approaches: studies that aim to address the same issue include distribution-based clustering (Preheim et al., 2013), cluster-free filtering (Tikhonov et al., 2015), Swarm (Mahé et al., 2015), and recently introduced DADA2 (Callahan et al., 2016).

Potential new directions for a more accurate depiction of microbial communities through marker gene amplicons come with new questions. What should microbial ecologists do with all the data they have generated, and plan to generate during the years to come? What are the computational and ecological issues that will need to be addressed for new methods to be more accessible in the field? Although our collection does not promise answers to these questions, we hope it will further stimulate the community of microbial ecologists and the developers of widely used software platforms to move beyond the use of OTUs that require arbitrary percent similarity cut-offs.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

AME was supported by the University of Chicago and the Marine Biological Laboratory collaboration award.

REFERENCES

- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., et al. (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* 4, 1111–1119. doi: 10.1111/2041-210X.12114
- Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., and Sogin, M. L. (2014). Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* 9, 968–979. doi: 10.1038/ismej.2014.195
- Eren, A. M., Sogin, M. L., Morrison, H. G., Vineis, J. H., Fisher, J. C., Newton, R. J., et al. (2015). A single genus in the gut microbiome reflects host preference and specificity. *ISME J.* 9, 90–100. doi: 10.1038/ismej.2014.97
- Jost, L. (2006). Entropy and diversity. *Oikos* 113, 363–375. doi: 10.1111/j.2006-0030-1299.14714.x
- Kleindienst, S., Seidel, M., Zierovogel, K., Grim, S., Loftis, K., Harrison, S., et al. (2015). Chemical dispersants can suppress the activity of natural oil-degrading microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14900–14905. doi: 10.1073/pnas.1507380112
- Koeppel, A. F., and Wu, M. (2013). Surprisingly extensive mixed phylogenetic and ecological signals among bacterial operational taxonomic units. *Nucleic Acids Res.* 41, 5175–5188. doi: 10.1093/nar/gkt241

- Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *Peer J.* 3:e1420. doi: 10.7717/peerj.1420
- Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E., and Borisy, G. G. (2016). Biogeography of a human oral microbiome at the micron scale. *Proc. Natl. Acad. Sci. U.S.A.* 113, 201522149. doi: 10.1073/pnas.1522149113
- McLellan, S. L., and Eren, A. M. (2014). Discovering new indicators of fecal pollution. *Trends Microbiol.* 22, 697–706. doi: 10.1016/j.tim.2014.08.002
- Nguyen, N.-P., Warnow, T., Pop, M., and White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *npj Biofilms Microbiomes* 2, 16004. doi: 10.1038/npjbiofilms.2016.4
- Preheim, S. P., Perrotta, A. R., Martin-Platero, A. M., Gupta, A., and Alm, E. J. (2013). Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl. Environ. Microbiol.* 79, 6593–6603. doi: 10.1128/AEM.00342-13
- Reveillaud, J., Maignien, L., Eren, A. M., Huber, J. A., Apprill, A., Sogin, M. L., et al. (2014). Host-specificity among abundant and rare taxa in the sponge microbiome. *ISME J.* 8, 1198–1209. doi: 10.1038/ismej.2013.227
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Technol. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Tikhonov, M., Leach, R. W., and Wingreen, N. S. (2015). Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J.* 9, 68–80. doi: 10.1038/ismej.2014.117

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Eren, Sogin and Maignien. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.