⬢ISME

High proportions of bacteria and archaea across most biomes remain uncultured

Andrew D. Steen [ID]¹,² · Alexander Crits-Christoph³ · Paul Carini [ID]⁴ · Kristen M. DeAngelis [ID]¹,⁵ · Noah Fierer [ID]²,⁶ · Karen G. Lloyd [ID]¹,² · J. Cameron Thrash [ID]⁷

Abstract

A recent paper by Martiny argues that "high proportions" of bacteria in diverse Earth environments have been cultured. Here we reanalyze a portion of the data in that paper, and argue that the conclusion is based on several technical errors, most notably a calculation of sequence similarity that does not account for sequence gaps, and the reliance on 16S rRNA gene amplicons that are known to be biased towards cultured organisms. We further argue that the paper is also based on a conceptual error: namely, that sequence similarity cannot be used to infer "culturability" because one cannot infer physiology from 16S rRNA gene sequences. Combined with other recent, more reliable studies, the evidence supports the conclusion that most bacterial and archaeal taxa remain uncultured.

What fraction of bacterial and archaeal cells on Earth have been cultured? This simple question has a complex answer. In 1932, Razumov showed that cultivation-based cell enumerations yield far smaller estimates of cell abundance than microscopic counts [1]. More recently, it

✉ Andrew D. Steen
asteen1@utk.edu

1 Department of Microbiology, University of Tennessee, Knoxville, TN 37996, USA

2 Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN 37996, USA

3 Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

4 Department of Environmental Science, BIO5 Institute, University of Arizona, Tucson, AZ, USA

5 Department of Microbiology, University of Massachusetts, Amherst, MA 01003, USA

6 Department of Ecology and Evolutionary Biology, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80305, USA

7 Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

has become clear that many microbial taxa are so distantly related to laboratory strains that they belong to entirely new families, orders, or even phyla with no cultured representatives [2, 3]. Most bacterial phyla lack cultured representatives [3]. Quantifying the fraction of cells on Earth that are represented in culture collections remains an important—but essential—analytical and conceptual challenge given that microbial physiology is an emergent property that is not easily inferred from sequence data alone.

In a recent Brief Communication, Martiny's answer to this question is summarized in the title, "High proportions of bacteria are culturable across major biomes", and his final sentence, "…many if not most abundant lineages across diverse biomes are already in culture" [4]. Here we argue that these conclusions are based on inadequate and inherently biased datasets and analyses as well as the mistaken logic that an organism is culturable if its 16S rRNA gene sequence is >97% similar to a laboratory culture. Below we highlight existing evidence that the majority of cells in most environments are only distantly related to organisms that have been cultured.

First, Martiny relies on full length, PCR-amplified 16S rRNA gene sequences to assess the abundance of microbial taxa in the environment. PCR primers commonly used to amplify 16S rRNA genes are biased toward certain taxa and completely miss newly discovered lineages, some of which can be identified from metagenomes [5–7].

Because shotgun sequencing of environmental DNA does not rely on sequence-specific PCR amplification prior to sequencing, it is not biased in the same way as amplification-based approaches. Thus, metagenomes are expected to contain 16S rRNA gene sequences in proportions that more closely approximate their absolute abundances in environmental DNA. A recent meta-analysis comparing 16S rRNA gene sequences retrieved from either amplicon or metagenomic datasets showed that commonly used primers targeting the 16S rRNA gene yielded a considerably higher abundance of cultured organisms across most publically available datasets [8]. While the magnitude of the difference between primer-based analyses and metagenomic-based analyses will differ depending on PCR primer selection and the environment in question, PCR-based assessments of culturability considerably and consistently over-represent the fraction of cultured organisms relative to metagenomic-based studies. To estimate the magnitude of this effect, we compared abundances of 16S rRNA gene sequences obtained by PCR amplification to abundances obtained from metagenomic libraries, in each of 14 broad environment types, using data originally reported in Lloyd et al. [8]. Sequence abundance in metagenomic libraries was measured as the median number of reads recruiting to each sequence, as reported by IMG/M [9], which represents the abundance of each 16S rRNA gene sequence in DNA extracted from each sample. We note that the PCR-amplified data and metagenomic data come from different studies, so they represent trends within environments rather than a head-to-head comparison of results from identical samples.

Second, Martiny used the SeqMatch tool provided by the Ribosomal Database Project (RDP) to identify sequence similarity between environmental 16S rRNA gene sequences and isolates [10]. SeqMatch reports a "similarity score" that does not include internal gapped alignment positions and only reports differences between aligned bases. This score is calculated differently from common identity scores between full length sequences, including default scores for programs such as BLAST, USEARCH, CD-HIT, UPARSE, and mothur, which count internal sequence insertions and deletions as biologically valid differences [11]. Because the length of the 16S rRNA genes can vary considerably [12], gaps are important markers of evolutionary distance [6]. The SeqMatch tool also reports sequence identity to the best hit in the RDP alignment, which does not align the most variable regions of the 16S rRNA gene [13], yet these regions are the most likely to contain substitutions between species. In addition, 1–2% of the sequences identified by SeqMatch as the closest isolate match to environmental sequences in the Martiny analysis are misidentified as isolates, and can be traced back to unidentified environmental sequences (Tables S1 and S2). These database errors would

have introduced a small additional bias towards over-estimating the abundance of cultured sequences. Lastly, the limitation of the analysis to only 100 sequences per sample is also likely to greatly underestimate the numbers of taxa identified from a given sample.

To demonstrate how database and analysis choice can impact results, we reanalyzed all sequences from Jangid et al. [14] and Santelli et al. [15] from the Martiny analysis, along with 48,000 representative OTU sequences obtained by Karst et al. [16] (Fig. 1). The Karst et al. data set was generated using novel methods based on physical concentration, size selection, and both short- and long-read sequencing of small subunit rRNA molecules, which is free from primer bias. We report sequence identities from the *align.seqs* command in the mothur package (counting internal, and not external, sequence gaps as biological differences) to both reported isolates and type strains in the SILVA [17] database (47,917 sequences), and a version of the RDP database filtered to contain only isolates based on RDP's "isolate filter" (513,426 sequences). We found lower median best hit percent sequence identities than those reported by Martiny, and often much lower percentages of sequences with >97% identity to an isolate in the dataset without primer biases than in the PCR studies. Only 6% of the representative OTU sequences reported by Karst et al. [16] were found to have >97% identity to any isolate in the larger filtered RDP dataset. Methods are described in more detail in the supplemental. Our revised results are closer to what was found for a large database, including nearly all published metagenomes, of which 2–18% of 16S rRNA gene sequences were >96.6% similar to a sequence from a cultured microbe [8]. This cutoff of 96.6% similarity is the upper 95% confidence interval of the median sequence identity within genera [16]. It is also worth noting that another study using the SILVA database and only full-length sequences concluded that "18.9% of bacterial sequences and 6.8% of archaeal sequences have come from isolated organisms", although there was great variation across phyla [18]. However, in no phylum were sequences from cultured organisms in the majority (Table 1).

Finally, we disagree on the conceptual linkage between "culturability" and 16S rRNA gene sequence similarity to a culture. The 97% 16S rRNA gene sequence similarity cutoff is frequently used to group microbes into species [19, 20], and it is true that microbes within this cutoff often share many traits that are important to culturability [21]. However, a great deal of genomic and phenotypic diversity can exist among strains of the same species [22, 23], with important implications for an organism's ecological function and the conditions required to culture it [24–27]. Physiological state can also influence culture yields: for instance, variable rates of dormancy can have profound impacts on assessments of "culturability" within a single
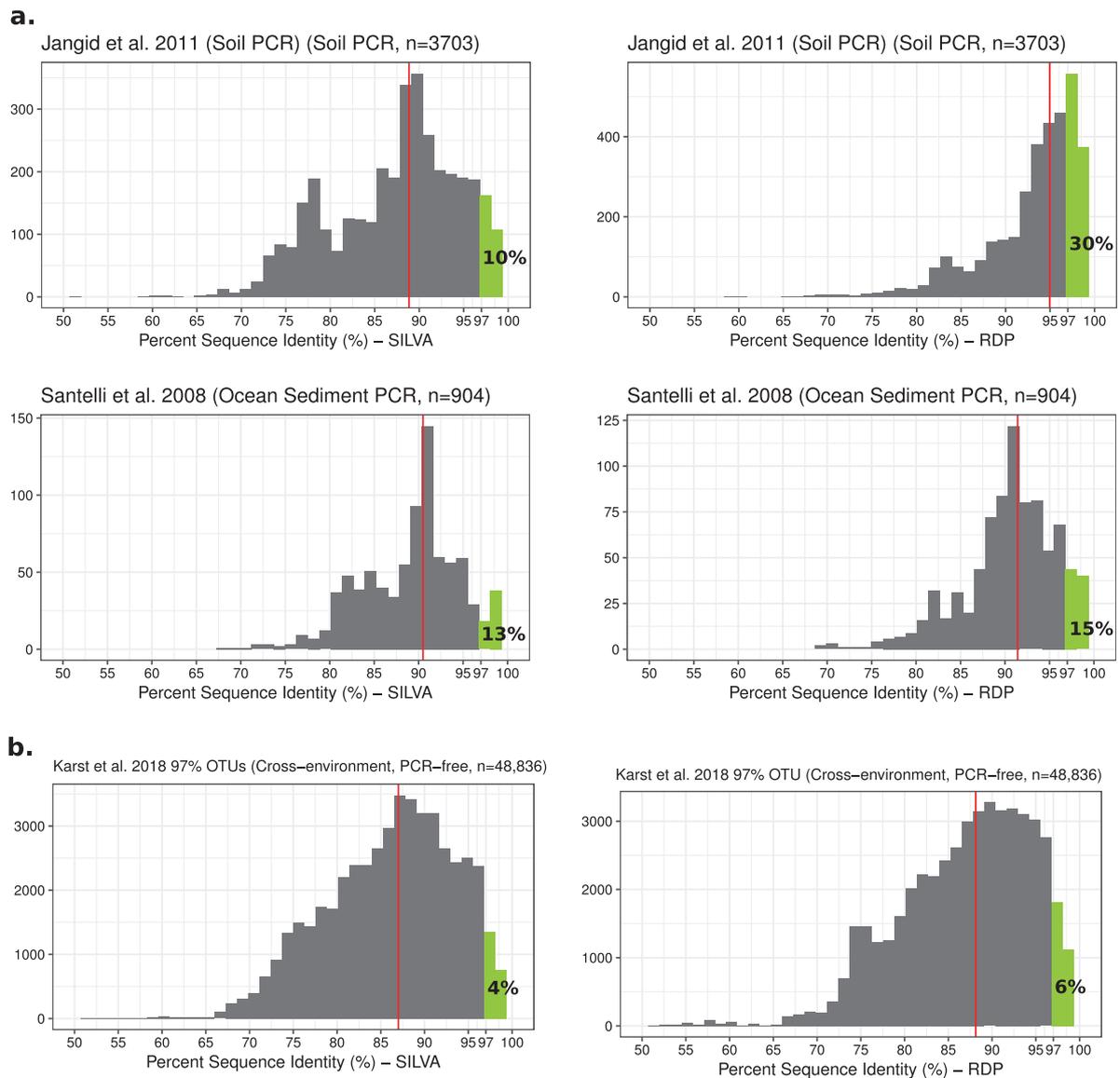
a.

b.

Fig. 1 Best hit percent identities of environmental sequences to both isolate 16S rRNA gene sequences in the SILVA database (left column) and a filtered version of the RDP database (right column) from **a** two PCR-based studies analyzed in Martiny 2019, and **b** a large set of dereplicated OTU sequences free from primer bias and from diverse environments [16]. Median best hit percent identities for each dataset are marked with a red line and the percentage of sequences, which had best hits at >97% identity are marked in green

Table 1 Enrichment of 16S rRNA gene sequences from organisms at different levels of taxonomic novelty due to PCR amplification, values are calculated as the ratio of the median fraction of 16S rRNA gene sequences in 4743 amplicon-based datasets to the median read depth of 16S rRNA gene sequences in 1504 unamplified metagenomes. (Data from figures 2 and 3 of Lloyd et al. [8])

Enrichment in amplicon libraries vs metagenome-derived sequences	Bacteria median (25th percentile, 75th percentile)	Archaea median (25th percentile, 75th percentile)
Cultured species (≥96.6% 16S rRNA gene similarity)	180% (140, 330%)	270% (180, 430%)
Uncultured genus to class	98% (76, 110%)	70% (55, 120%)
Uncultured phylum	54% (32, 82%)	70% (46, 94%)

population [28]. Thus, it is impossible to reliably assess whether an organism is "culturable" based on 16S rRNA gene sequence alone.

We agree with Martiny that the oft-repeated axiom that only 1% of microbial cells in all environments are culturable should be retired. As Martiny clearly lays out, the precise meaning of this statement is often nebulous, and the underlying data are hard to identify. One recent meta-analysis found that a median of 0.5% of cells identified in direct counts could be grown in culture using standard techniques [8], but the variance was large (interquartile range from 0.025 to 4.3%; Fig S1). Individual studies have found culturability yields of 70% or more in environments as diverse as the human gut, surface marine sediments, and lakes in which volcanic ash had recently been deposited [29–31]. Furthermore, the term "culturable" only makes sense in the context of specified culturing conditions, a distinction that Martiny draws in the difference between H1 (that 1% of cells can be cultured) and H3 (that 1% of cells will grow on a standard agar plate). Culturing efforts to characterize microbial communities should rely on diverse methods and media to maximize yield [32] and methodological innovations have opened the door to culturing previously uncultured taxa [33–37].

Thus, we argue that it is impossible to know whether a microbe is culturable until it has been cultured, so the term "unculturable" should be avoided in favor of "uncultured." Historically, the cultivation and study of microbes in the laboratory has been integral to our understanding of the microbial roles in the world around us, and even in the age of bioinformatics it remains essential. In order to understand the degree to which we have characterized the microbial world, we need to measure the fraction of microbial cells on Earth that share physiologies with cultured cells. Getting this estimate right is an important academic question, but it also has implications for allocation of resources by funding agencies. An overestimate of the degree to which culture collections reflect the Earth's microbiome would discourage funders from supporting efforts to culture important, uncultured taxa. We hope that future culturing efforts will be successful enough that most bacteria and archaea from most environments will have representation in culture collections. At present, however, the best evidence indicates that this is very far from the case.

Compliance with ethical standards

References

1. Razumov AS. The direct method of calculation of bacteria in water: comparison with the Koch method. Mikrobiologija. 1932;1:131–46.
2. Dick GJ, Baker BJ. Omic Approaches in Microbial Ecology: Charting the Unknown: analysis of whole-community sequence data is unveiling the diversity and function of specific microbial groups within uncultured phyla and across entire microbial ecosystems. Microbe Mag. 2013;8:353–60.
3. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016;1:16048.
4. Martiny AC High proportions of bacteria are culturable across major biomes. ISME J. 2019;4:2125–8.
5. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol. 1998;64:3724–30.
6. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 2015;523:208–11.
7. Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol. 2016;1:15032.
8. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems 2018;3:e00055–18.
9. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res 2019;47:D666–77.
10. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42:D633–42.
11. Flynn JM, Brown EA, Chain FJJ, MacIsaac HJ, Cristescu ME. Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. Ecol Evol. 2015;5:2252–66.
12. Smit S, Widmann J, Knight R. Evolutionary rates vary among rRNA structural elements. Nucleic Acids Res. 2007;35:3339–54.
13. Schloss PD. A high-throughput DNA sequence aligner for microbial ecology studies. PLoS ONE. 2009;4:e8230.

14. Jangid K, Williams MA, Franzluebbers AJ, Schmidt TM, Coleman DC, Whitman WB. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. Soil Biol Biochem. 2011;43:2184–93.

15. Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML, et al. Abundance and diversity of microbial life in ocean crust. Nature. 2008;453:653–6.

16. Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat Biotechnol. 2018;36:190–5.

17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.

18. Schloss PD, Girard RA, Martin T, Edwards J, Thrash JC. Status of the archaeal and bacterial census: an update. MBio 2016;7: e00201–16.

19. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol. 2014;12:635–45.

20. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci USA. 2005;102:2567–72.

21. Oberhardt MA, Zarecki R, Gronow S, Lang E, Klenk H-P, Gophna U, et al. Harnessing the landscape of microbial culture media to predict new organismmedia pairings. Nat Commun. 2015;6:8493.

22. Mira A, Martín-Cuadrado AB, D'Auria G, Rodríguez-Valera F. The bacterial pangenome:a new paradigm in microbiology. Int Microbiol. 2010;13:45–57.

23. Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT. Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. Proc Natl Acad Sci USA. 2011;108:7200–5.

24. Schloter M, Lebuhn M, Heulin T, Hartmann A. Ecology and evolution of bacterial microdiversity. FEMS Microbiol Rev. 2000;24:647–60.

25. Larkin AA, Martiny AC. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environ Microbiol Rep. 2017;9:55–70.

26. Chase AB, Karaoz U, Brodie EL, Gomez-Lunar Z, Martiny AC, Martiny JBH. Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. MBio 2017;8:e01809–17.

27. Jaspers E, Overmann J. Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. Appl Environ Microbiol. 2004;70:4831–9.

28. Buerger S, Spoering A, Gavrish E, Leslin C, Ling L, Epstein SS. Microbial scout hypothesis, stochastic exit from dormancy, and the nature of slow growers. Appl Environ Microbiol. 2012; 78:3221–8.

29. Staley JT, Konopka A. Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. Ann Rev Microbiol 1985;39:321–46.

30. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature. 2016; 533:543–6.

31. Kaeberlein T, Lewis K, Epstein SS. Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. Science. 2002;296:1127–9.

32. Pold G, Billings AF, Blanchard JL, Burkhardt DB, Frey SD, Melillo JM, et al. Longterm warming alters carbohydrate degradation potential in temperate forest soils. Appl Environ Microbiol. 2016;82:6518–30.

33. Kato S, Yamagishi A, Daimon S, Kawasaki K, Tamaki H, Kitagawa W, et al. Isolation of Previously Uncultured Slow-Growing Bacteria by Using a Simple Modification in the Preparation of Agar Media. Appl Environ Microbiol 2018;84:e00807–18.

34. Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH. Are uncultivated bacteria really uncultivable? Microbes Environ. 2012;27:356–66.

35. Berdy B, Spoering AL, Ling LL, Epstein SS. In situ cultivation of previously uncultivable microorganisms using the ichip. Nat Protoc. 2017;12:2232–42.

36. Tanaka T, Kawasaki K, Daimon S, Kitagawa W, Yamamoto K, Tamaki H, et al. A hidden pitfall in the preparation of agar media undermines microorganism cultivability. Appl Environ Microbiol. 2014;80:7659–66.

37. Zengler K, Toledo G, Rappe M, Elkins J, Mathur EJ, Short JM, et al. Cultivating the uncultured. Proc Natl Acad Sci USA. 2002;99:15681–6.