UC Riverside
UC Riverside Previously Published Works

Title
Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units.

Permalink
https://escholarship.org/uc/item/45n7t2w1

Journal
mSphere, 3(4)

ISSN
2379-5042

Authors
Glassman, Sydney I
Martiny, Jennifer BH

Publication Date
2018-07-01

DOI
10.1128/msphere.00148-18

Peer reviewed

Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units

Sydney I. Glassman,[a,b] Jennifer B. H. Martiny[a]

[a]Department of Ecology and Evolutionary Biology, University of California—Irvine, Irvine, California, USA
[b]Department of Microbiology and Plant Pathology, University of California—Riverside, Riverside, California, USA

ABSTRACT Recent discussion focuses on the best method for delineating microbial taxa, based on either exact sequence variants (ESVs) or traditional operational taxonomic units (OTUs) of marker gene sequences. We sought to test if the binning approach (ESVs versus 97% OTUs) affected the ecological conclusions of a large field study. The data set included sequences targeting all bacteria (16S rRNA) and fungi (internal transcribed spacer [ITS]), across multiple environments diverging markedly in abiotic conditions, over three collection times. Despite quantitative differences in microbial richness, we found that all α and β diversity metrics were highly positively correlated ($r > 0.90$) between samples analyzed with both approaches. Moreover, the community composition of the dominant taxa did not vary between approaches. Consequently, statistical inferences were nearly indistinguishable. Furthermore, ESVs only moderately increased the genetic resolution of fungal and bacterial diversity (1.3 and 2.1 times OTU richness, respectively). We conclude that for broadscale (e.g., all bacteria or all fungi) α and β diversity analyses, ESV or OTU methods will often reveal similar ecological results. Thus, while there are good reasons to employ ESVs, we need not question the validity of results based on OTUs.

IMPORTANCE Microbial ecologists have made exceptional improvements in our understanding of microbiomes in the last decade due to breakthroughs in sequencing technologies. These advances have wide-ranging implications for fields ranging from agriculture to human health. Due to limitations in databases, the majority of microbial ecology studies use a binning approach to approximate taxonomy based on DNA sequence similarity. There remains extensive debate on the best way to bin and approximate this taxonomy. Here we examine two popular approaches using a large field-based data set examining both bacteria and fungi and conclude that there are not major differences in the ecological outcomes. Thus, it appears that standard microbial community analyses are not overly sensitive to the particulars of binning approaches.

KEYWORDS Illumina MiSeq, bacteria, exact sequence variants (ESVs), fungi, microbial ecology, operational taxonomic units (OTUs)

Characterization of microbial communities by amplicon sequencing introduces biases and errors at every step. Hence, choices concerning all aspects of molecular processing from DNA extraction method (1) to sequencing platform (2) are debated. Further downstream, the choices for computational processing of amplicon sequences are similarly deliberated (e.g., see references 3 to 5). Yet despite these ongoing debates, microbial ecology has made great strides toward characterizing and testing hypotheses in environmental and host-associated microbiomes (e.g., see references 6 and 7).

Within microbiome studies, operational taxonomic units (OTUs) have been used to delineate microbial taxa, as the majority of microbial diversity remains unrepresented

in global databases (8). While any degree of sequence similarity could be used to denote individual taxa, a 97% sequence similarity cutoff became standard within microbial community analyses. This cutoff attempted to balance previous standards for defining microbial species (9) and recognition of spurious diversity accumulated through PCR and sequencing errors (10, 11).

Recently, it has been suggested that taxa should be defined based on exact nucleotide sequences of marker genes. Delineation of taxa by exact sequence variants (ESVs), also termed amplicon sequence variants (ASVs [12]) or zero-radius OTUs (zOTUs [13]), is not only expected to increase taxonomic resolution, but could also simplify comparisons across studies by eliminating the need for rebinning taxa when data sets are merged. Due to these advantages, there has been a surge in bioinformatic pipelines that seek to utilize ESVs and minimize specious sequence diversity (13–15). Moreover, some proponents have stated that ESVs should replace OTUs altogether (12). However, as with the adoption of any new approach, there remains a need to quantify how this new method compares to a large body of previous research. Furthermore, OTU classifications remain biologically useful for comparing diversity across large data sets (7) or identifying clades that share traits (16).

Here, we tested if use of ESVs versus 97% OTUs affected the ecological conclusions, including treatment effects and α and β diversity patterns, from a large field study of leaf litter communities. This study included a "site" and "inoculum" treatment, in which all microbial communities were reciprocally transplanted into all five sites (see Text S1 in the supplemental material) along an elevation gradient (17). We sequenced both bacteria (16S rRNA) and fungi (internal transcribed spacer 2 [ITS2]) from litterbags collected at three time points (6, 12, and 18 months after deployment) in separate sequencing runs. While we expected that the binning approach would alter observed richness, we hypothesized that it might not alter trends in α and β diversity, but that these results might differ based on the amplicon sequenced.

In total, we analyzed >15 million bacterial and >20 million fungal sequences using UPARSE v10 (see Table S1 in the supplemental material), which allowed for a direct comparison of ESV versus 97% OTU approaches by keeping all other aspects of quality filtering and merging consistent (4). We selected a direct comparison with 97% OTUs as it is the most standard threshold and the clustering algorithms appear to be most effective at this level (R. Edgar, personal communication). A recent study also found that clustering thresholds from 87% to 99% yield highly stable results (18).

ESV and OTU α diversity was strongly correlated across samples using four metrics for both bacteria and fungi (mean Pearson's $r = 0.95 \pm 0.02$; all P values are <0.001). For three metrics (Berger-Parker, Shannon, and Simpson), the ESV and OTU approaches were not only highly correlated (mean Pearson's $r = 0.95 \pm 0.02$), but nearly equivalent in their values (mean slope = 0.97) (see Table S2 in the supplemental material). For observed richness, ESV versus OTU was also highly correlated across all time points/ sequencing runs (Pearson's $r > 0.92$) (Fig. 1A and B). However, bacterial OTU richness was approximately half of ESV richness for the same sample (mean slope = 0.46), and fungal OTU richness was approximately three-quarters of ESV richness (mean slope = 0.79). We speculate that this difference between bacteria and fungi is due to the coarser phylogenetic breadth of the 16S versus ITS genetic regions.

β diversity metrics were also strongly correlated across samples for ESVs and OTUs (Bray-Curtis average Mantel's $r = 0.96$ for bacteria and 0.98 for fungi; all P values are <0.01 [Fig. 1C and D]), whether assessed by abundance-based (Bray-Curtis) or presence-absence (Jaccard) metrics (Table S2). Moreover, the values of the β diversity metrics were nearly identical regardless of binning approach (slopes of ~1).

The highly correlated α and β diversity metrics indicated that results based on these metrics should yield similar ecological conclusions. Indeed, the patterns of bacterial and fungal richness and community composition across the elevation gradient were nearly indistinguishable (Fig. 2; see Fig. S1 in the supplemental material), as were the statistical tests for both richness (see Table S3 in the supplemental material) and community composition (see Tables S4 and S5 in the supplemental material). Moreover, family- and
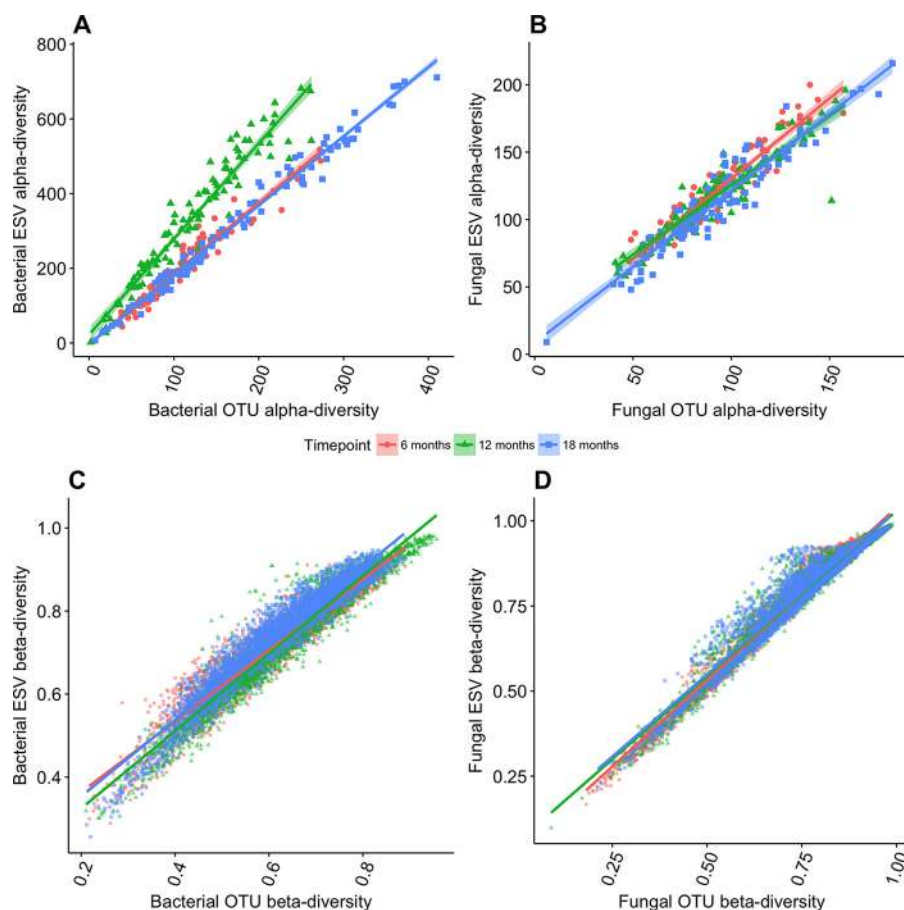
FIG 1 (A and B) Comparison of observed α diversity for (A) bacteria and (B) fungi as assayed by the richness of 97% similar operational taxonomic units (OTUs) versus exact sequence variants (ESVs). Numbers are total observed richness after normalizing to 10,000 sequences per sample from three time points (16, 12, and 18 months). (C and D) Comparison of observed β diversity for (C) bacteria and (D) fungi as assayed by the Bray-Curtis dissimilarity for OTUs versus ESVs from three time points (16, 12, 18 months).

genus-level compositions at each site along the gradient were virtually identical for bacteria (see Fig. S2 in the supplemental material) and highly similar for fungi (see Fig. S3 in the supplemental material), with no taxa being over- or underrepresented in the ESV versus OTU approaches for bacteria (Fig. S2C) and only one for fungi (Fig. S3C). We also included a mock community of eight distinct bacterial species in our PCR and sequencing runs. Both approaches resulted in highly similar mock community composition (see Fig. S4 in the supplemental material). Thus, we found no evidence that ESVs yield better taxonomic resolution or are more sensitive to detecting treatment effects (12). If anything, the ESV method appeared to be slightly less sensitive to detecting treatment effects on richness than the OTU method, especially for fungi in which fewer significant treatment effects were detected using ESVs (Table S3).

Despite quantitative differences in microbial richness, ecological interpretation of our large bacterial and fungal community data set was robust to the use of ESVs versus 97% OTUs. Thus, even though there are good reasons to take an ESV approach, we need not question the validity of ecological results based on OTUs. Indeed, while previous studies have found that ESVs can help explain additional variation among samples (19, 20), the α and β diversity patterns of ESVs and OTUs in these studies were also quite similar. In general, we suspect that the robustness of such comparisons will vary depending on the breadth of the microbial community targeted. For instance, here we characterized all bacteria and fungi in a diverse environmental community, as opposed to a narrower subset of taxa or a less diverse, host-associated community.
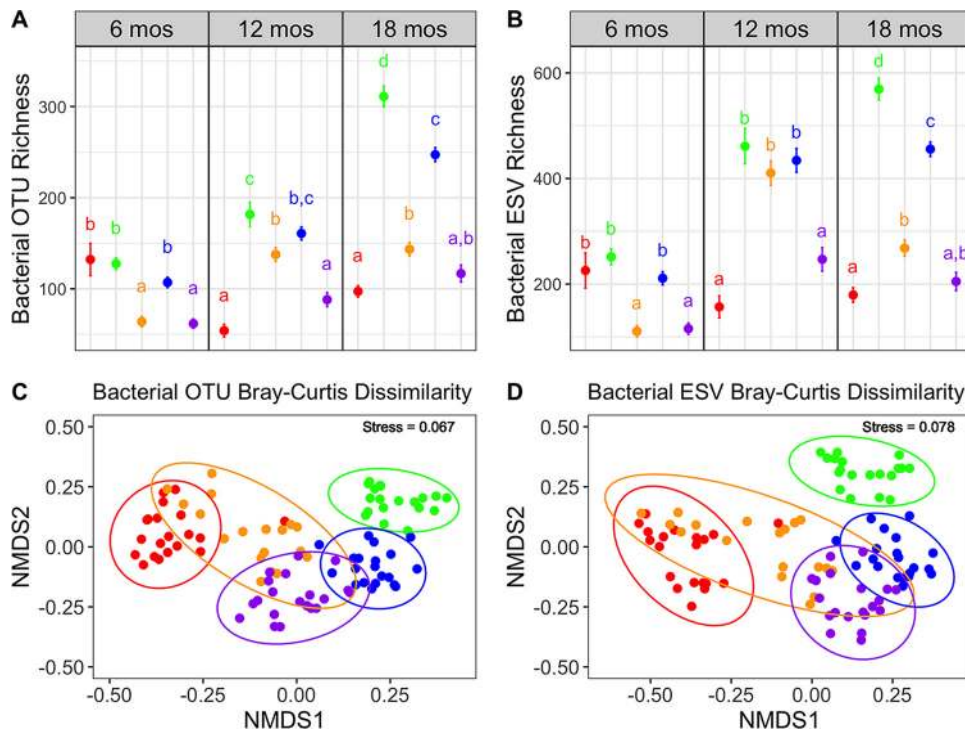
FIG 2 (A and B) Comparison of α diversity results using (A) operational taxonomic units (OTUs) versus (B) exact sequencing variants (ESVs) for bacteria across the elevation gradient at three time points (16, 12, and 18 months). Each point represents mean observed richness per litterbag per site, and lines indicated standard error (averaged across five inoculum treatments and four replicates; $n = 20$). Letters represent Tukey's honestly significant difference (HSD) test significant differences across sites within a time point. (C and D) Comparison of β diversity results using nonmetric multidimensional scaling (NMDS) ordination of Bray-Curtis community dissimilarity of (C) bacterial OTUs and (D) bacterial ESVs colored by site at the final time point (18 months). Ellipses represent 95% confidence intervals around the centroid. Colors represent sites along the elevation gradient ranging from the lowest elevation (red = 275 m) to highest elevation (purple = 2,240 m), with middle elevation sites colored as follows: green = 470 m, orange = 1,280 m, and blue = 1,710 m.

Finally, both 97% OTUs and ESVs mask ecologically important trait variation of individual taxa (19, 21). In our study, ESVs only slightly increased the detection of fungal and bacterial diversity (1.3 and 2.1 times OTU richness, respectively), highlighting that ribosomal marker genes at any resolution are generally poor targets for improving genetic resolution within a microbial community. For example, it is widely known that many taxa can share the same 16S rRNA (21) or ITS (22). Thus, if strain identification is critical, then a full genome (21) or amplicon of a less conservative marker gene (23) is required. However, for broadscale community α and β diversity patterns, although the vagaries of molecular and bioinformatics processing inevitably add noise to microbial sequencing data, strong community-level signals will likely emerge with suitable study designs and statistics regardless of binning approach.

Data availability. Sequences were submitted to the National Center for Biotechnology Information Sequence Read Archive under accession no. SRP150375 and BioProject no. PRJNA474008. All data and scripts to recreate all figures and statistics from this article can be found on github at https://github.com/sydneyg/OTUvESV.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSphere.00148-18.

TEXT S1, PDF file, 0.1 MB.
FIG S1, DOCX file, 0.1 MB.
FIG S2, DOCX file, 0.1 MB.
FIG S3, DOCX file, 0.1 MB.

FIG S4, DOCX file, 0.03 MB.
TABLE S1, DOCX file, 0.02 MB.
TABLE S2, DOCX file, 0.1 MB.
TABLE S3, DOCX file, 0.03 MB.
TABLE S4, DOCX file, 0.03 MB.
TABLE S5, DOCX file, 0.03 MB.

REFERENCES

1. Frostegård A, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P. 1999. Quantification of bias related to the extraction of DNA directly from soils. Appl Environ Microbiol 65:5409–5420.

2. Claesson MJ, Wang QO, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW. 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res 38:e200. https://doi.org/10.1093/nar/gkq873.

3. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.

4. Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998. https://doi.org/10.1038/nmeth.2604.

5. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.

6. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463. https://doi.org/10.1038/nature24621.

7. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. Science 359:320–325. https://doi.org/10.1126/science.aap9516.

8. Moyer CL, Dobbs FC, Karl DM. 1994. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl Environ Microbiol 60:871–879.

9. Stackebrandt E, Goebel BM. 1994. A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol 44:846–849.

10. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol 12:118–123. https://doi.org/10.1111/j.1462-2920.2009.02051.x.

11. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71:8966–8969. https://doi.org/10.1128/AEM.71.12.8966-8969.2005.

12. Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11:2639–2643. https://doi.org/10.1038/ismej.2017.119.

13. Edgar RC. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon reads. bioRxiv https://doi.org/10.1101/081257.

14. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

15. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu ZZ, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2:e00191-16. https://doi.org/10.1128/mSystems.00191-16.

16. Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. 2009. Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. Environ Microbiol 11:823–832. https://doi.org/10.1111/j.1462-2920.2008.01803.x.

17. Baker NR, Allison SD. 2017. Extracellular enzyme kinetics and thermodynamics along a climate gradient in southern California. Soil Biol Biochem 114:82–92. https://doi.org/10.1016/j.soilbio.2017.07.005.

18. Botnen SS, Davey ML, Halvorsen R, Kauserud H. 2018. Sequence clustering threshold has little effect on the recovery of microbial community structure. Mol Ecol Resour https://doi.org/10.1111/1755-0998.12894.

19. Needham DM, Sachdeva R, Fuhrman JA. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. ISME J 11:1614–1629. https://doi.org/10.1038/ismej.2017.29.

20. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. 2015. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J 9:968–979. https://doi.org/10.1038/ismej.2014.195.

21. Chase AB, Karaoz U, Brodie EL, Gomez-Lunar Z, Martiny AC, Martiny JBH. 2017. Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. mBio 8:e01809-17. https://doi.org/10.1128/mBio.01809-17.

22. Dettman JR, Jacobson DJ, Taylor JW. 2006. Multilocus sequence data reveal extensive phylogenetic species diversity within the Neurospora discreta complex. Mycologia 98:436–446. https://doi.org/10.3852/mycologia.98.3.436.

23. Larkin AA, Martiny AC. 2017. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environ Microbiol Rep 9:55–70. https://doi.org/10.1111/1758-2229.12523.