PLOS BIOLOGY

FORMAL COMMENT

Response to "Vast (but avoidable) underestimation of global biodiversity"

Stilianos Louca[1,2]*, **Florent Mazel**[3,4], **Michael Doebeli**[3,5,6], **Laura Wegener Parfrey**[3,4,5]

1 Department of Biology, University of Oregon, Eugene, Oregon, United States of America, **2** Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon, United States of America, **3** Biodiversity Research Centre, University of British Columbia, Vancouver, Canada, **4** Department of Botany, University of British Columbia, Vancouver, Canada, **5** Department of Zoology, University of British Columbia, Vancouver, Canada, **6** Department of Mathematics, University of British Columbia, Vancouver, Canada

* louca.research@gmail.com

Citation: Louca S, Mazel F, Doebeli M, Parfrey LW (2021) Response to "Vast (but avoidable) underestimation of global biodiversity". PLoS Biol 19(8): e3001362. https://doi.org/10.1371/journal.pbio.3001362

Received: March 25, 2021

Accepted: July 13, 2021

Published: August 13, 2021

Data Availability Statement: The R code used to estimate the microbiome overlap between animal samples is provided as Supplemental File 2 and is also available at: https://github.com/FloMazel/Overlap_Animal_Gut_Microbiota. The input ASVtables, provided by the Earth Microbiome Project, are publicly available and were downloaded from the Qiita website (https://qiita.ucsd.edu/study/description/11166) with Qiita biom table IDs being 93862, 93855, 93819, 93914, 93900, 93846, 93851 and 94483. Metadata used are available for download as supplemental material of Song al colleagues 2020 (S1 Data SET) (https://mbio.asm.org/content/mbio/11/1/e02901-19/DC1/embed/inline-supplementary-material-1.xlsx?download=

We welcome Wiens' efforts to estimate global animal-associated bacterial richness and thank him for highlighting points of confusion and potential caveats in our previous work on the topic [1]. We find Wiens' ideas worthy of consideration, as most of them represent a step in the right direction, and we encourage lively scientific discourse for the advancement of knowledge. Time will ultimately reveal which estimates, and underlying assumptions, came closest to the true bacterial richness; we are excited and confident that this will happen in the near future thanks to rapidly increasing sequencing capabilities. Here, we provide some clarifications on our work, its relation to Wiens' estimates, and the current status of the field.

First, Wiens states that we excluded animal-associated bacterial species in our global estimates. However, thousands of animal-associated samples were included in our analysis, and this was clearly stated in our main text (second paragraph on page 3).

Second, Wiens' commentary focuses on "S1 Text" of our paper [1], which was rather peripheral, and, hence, in the Supporting information. S1 Text [1] critically evaluated the rationale underlying previous estimates of global bacterial operational taxonomic unit (OTU) richness by Larsen and colleagues [2], but the results of S1 Text [1] did not in any way flow into the analyses presented in our main article. Indeed, our estimates of global bacterial (and archaeal) richness, discussed in our main article, are based on 7 alternative well-established estimation methods founded on concrete statistical models, each developed specifically for richness estimates from multiple survey data. We applied these methods to >34,000 samples from >490 studies including from, but not restricted to, animal microbiomes, to arrive at our global estimates, independently of the discussion in S1 Text [1].

Third, Wiens' commentary can yield the impression that we proposed that there are only 40,100 animal-associated bacterial OTUs and that *Cephalotes* in particular only have 40 associated bacterial OTUs. However, these numbers, mentioned in our S1 Text [1], were not meant to be taken as proposed point estimates for animal-associated OTU richness, and we believe that this was clear from our text. Instead, these numbers were meant as examples to demonstrate how strongly the estimates of animal-associated bacterial richness by Larsen and colleagues [2] would decrease simply by (a) using better justified mathematical formulas, i.e., with the same input data as used by Larsen and colleagues [2] but founded on an actual statistical model; (b) accounting for even minor overlaps in the OTUs associated with different animal genera; and/or (c) using alternative animal diversity estimates published by others [3], rather than those proposed by Larsen and colleagues [2]. Specifically, regarding (b), Larsen and colleagues [2] (pages 233 and 259) performed pairwise host species comparisons within various

true). Please also see the README file included with the code.

Funding: SL was supported by a startup grant by the University of Oregon. MD was supported by NSERC Discovery Grant nr. 219930. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

insect genera (for example, within the *Cephalotes*) to estimate on average how many bacterial OTUs were unique to each host species, then multiplied that estimate with their estimated number of animal species to determine the global animal-associated bacterial richness. However, since their pairwise host species comparisons were restricted to congeneric species, their estimated number of unique OTUs per host species does not account for potential overlaps between different host genera. Indeed, even if an OTU is only found "in one" *Cephalotes* species, it might not be truly unique to that host species if it is also present in members of other host genera. To clarify, we did not claim that all animal genera can share bacterial OTUs, but instead considered the implications of some average microbiome overlap (some animal genera might share no bacteria, and other genera might share a lot). The average microbiome overlap of 0.1% (when clustering bacterial 16S sequences into OTUs at 97% similarity) between animal genera used in our illustrative example in S1 Text [1] is of course speculative, but it is not unreasonable (see our next point). A zero overlap (implicitly assumed by Larsen and colleagues [2]) is almost certainly wrong. One goal of our S1 Text [1] was to point out the dramatic effects of such overlaps on animal-associated bacterial richness estimates using "basic" mathematical arguments.

Fourth, Wiens' commentary could yield the impression that existing data are able to tell us with sufficient certainty when a bacterial OTU is "unique" to a specific animal taxon. However, so far, the microbiomes of only a minuscule fraction of animal species have been surveyed. One can thus certainly not exclude the possibility that many bacterial OTUs currently thought to be "unique" to a certain animal taxon are eventually also found in other (potentially distantly related) animal taxa, for example, due to similar host diets and or environmental conditions [4–7]. As a case in point, many bacteria in herbivorous fish guts were found to be closely related to bacteria in mammals [8], and Song and colleagues [6] report that bat microbiomes closely resemble those of birds. The gut microbiome of caterpillars consists mostly of dietary and environmental bacteria and is not species specific [4]. Even in animal taxa with characteristic microbiota, there is a documented overlap across host species and genera. For example, there are a small number of bacteria consistently and specifically associated with bees, but these are found across bee genera at the level of the 99.5% similar 16S rRNA OTUs [5]. To further illustrate that an average microbiome overlap between animal taxa at least as large as the one considered in our S1 Text (0.1%) [1] is not unreasonable, we analyzed 16S rRNA sequences from the Earth Microbiome Project [6,9] and measured the overlap of microbiota originating from individuals of different animal taxa. We found that, on average, 2 individuals from different host classes (e.g., 1 mammalian and 1 avian sample) share 1.26% of their OTUs (16S clustered at 100% similarity), and 2 individuals from different host genera belonging to the same class (e.g., 2 mammalian samples) share 2.84% of their OTUs (methods in S1 Text of this response). A coarser OTU threshold (e.g., 97% similarity, considered in our original paper [1]) would further increase these average overlaps. While less is known about insect microbiomes, there is currently little reason to expect a drastically different picture there, and, as explained in our S1 Text [1], even a small average microbiome overlap of 0.1% between host genera would strongly limit total bacterial richness estimates. The fact that the accumulation curve of detected bacterial OTUs over sampled insect species does not yet strongly level off says little about where the accumulation curve would asymptotically converge; rigorous statistical methods, such as the ones used for our global estimates [1], would be needed to estimate this asymptote.

Lastly, we stress that while the present conversation (including previous estimates by Louca and colleagues [1], Larsen and colleagues [2], Locey and colleagues [10], Wiens' commentary, and this response) focuses on 16S rRNA OTUs, it may well be that at finer phylogenetic resolutions, e.g., at bacterial strain level, host specificity and bacterial richness are substantially

higher. In particular, future whole-genome sequencing surveys may well reveal the existence of far more genomic clusters and ecotypes than 16S-based OTUs.

Supporting information

S1 Text. Methods details on computing microbiome overlaps between animals.
(PDF)

S1 Data. Host taxonomies considered.
(CSV)

S2 Data. R code for computing microbiome overlaps.
(ZIP)

S3 Data. Computed Jaccard similarities.
(CSV)

Author Contributions

Conceptualization: Stilianos Louca, Florent Mazel, Michael Doebeli, Laura Wegener Parfrey.

Data curation: Florent Mazel.

Formal analysis: Florent Mazel.

Investigation: Florent Mazel.

Methodology: Florent Mazel.

Writing – original draft: Stilianos Louca, Florent Mazel, Michael Doebeli, Laura Wegener Parfrey.

Writing – review & editing: Stilianos Louca, Florent Mazel, Michael Doebeli, Laura Wegener Parfrey.

References

1. Louca S, Mazel F, Doebeli M, Parfrey WL. A census-based estimate of Earth's bacterial and archaeal diversity. PLoS Biol. 2019; 17(2):e3000106. https://doi.org/10.1371/journal.pbio.3000106 PMID: 30716065
2. Larsen BB, Miller EC, Rhodes MK, Wiens JJ. Inordinate fondness multiplied and redistributed: the number of species on Earth and the new pie of life. Q Rev Biol. 2017; 92(3):229–65. https://doi.org/10.1086/693564
3. Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. How many species are there on Earth and in the ocean? PLoS Biol. 2011; 9(8):e1001127. https://doi.org/10.1371/journal.pbio.1001127 PMID: 21886479
4. Hammer TJ, Janzen DH, Hallwachs W, Jaffe SP, Fierer N. Caterpillars lack a resident gut microbiome. Proc Natl Acad Sci U S A. 2017; 114(36):9641–6. https://doi.org/10.1073/pnas.1707186114 PMID: 28830993
5. Kwong WK, Medina LA, Koch H, Sing KW, Soh EJY, Ascher JS, et al. Dynamic microbiome evolution in social bees. Sci Adv. 2017; 3(3). https://doi.org/10.1126/sciadv.1600513 PMID: 28435856
6. Song SJ, Sanders JG, Delsuc F, Metcalf J, Amato K, Taylor MW, et al. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. mBio. 2020; 11(1). https://doi.org/10.1128/mBio.02901-19 PMID: 31911491
7. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science. 2011; 332(6032):970–4. https://doi.org/10.1126/science.1198719 PMID: 21596990

8. Sullam KE, Essinger SD, Lozupone CA, O'connor MP, Rosen GL, Knight R, et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Mol Ecol. 2012; 21 (13):3363–78. https://doi.org/10.1111/j.1365-294X.2012.05552.x PMID: 22486918

9. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017; 551:457–63. https://doi.org/10.1038/nature24621 PMID: 29088705

10. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proc Natl Acad Sci U S A. 2016; 113(21):5970–5. https://doi.org/10.1073/pnas.1521291113 PMID: 27140646

Response to: Vast (but avoidable) underestimation of global biodiversity
- Supplementary Material -

Stilianos Louca[1,2,*], Florent Mazel[3,4], Michael Doebeli[3,5,6] & Laura Wegener Parfrey[3,5,5]

[1]*Department of Biology, University of Oregon, Eugene, Oregon, USA*
[2]*Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon, USA*
[3]*Biodiversity Research Centre, University of British Columbia, Vancouver, Canada*
[4]*Department of Botany, University of British Columbia, Vancouver, Canada*
[5]*Department of Zoology, University of British Columbia, Vancouver, Canada*
[6]*Department of Mathematics, University of British Columbia, Vancouver, Canada*
[*]Corresponding author

S.1 Computing microbiome overlaps between animals

Here we describe our analysis of prokaryotic OTU overlap between samples of animal gut microbiota, originating from individual animals of different genera. We used the 16S rRNA sequencing data set described by Song et al. [1]. This data set updates the Earth Microbiome Project data set (release #1) [2] with additional samples from vertebrate hosts for a total of 460 host species in 374 genera, mostly of birds and mammals. We provide a summary of the host diversity that was published in Song et al. [1] and that we used here in Supplemental File 1.

Details concerning the sampling, DNA extraction, PCR, sequencing and processing can be found in Song et al. [1] and follow the Earth Microbiome Project protocol [2]. Briefly, the V4 region of the 16S rRNA was amplified using the 515f/806r EMP primers and amplicons were sequenced on Illumina (MiSeq and HiSeq) platforms. Because the data set by Song et al. [1] originates from several studies, each respective data set was processed in the same way to produce an OTU table. Here, we did not used the raw fastq files but instead used the processed OTU tables produced by Song et al. [1] from the reads trimmed at 90pb. Those table are provided on the Qiita website (see details below). In the pipeline by Song et al. [1], reads were quality-filtered, checked against the Greengenes database [3] to remove artifactual sequences, and sequencing errors were removed using the deblur algorithm [4], yielding amplicon sequence variants (ASVs, i.e. 16S sequences clustered at 100% sequence similarity). The resulting ASV tables are publicly available and were downloaded from the Qiita website (`https://qiita.ucsd.edu/study/description/11166`) with Qiita biom table IDs being 93862,93855,93819,93914,93900,93846,93851 and 94483. Metadata are available for download as supplemental material of Song et al. [1] (DATA SET S1 therein) (`https://mbio.asm.org/content/mbio/11/1/e02901-19/DC1/embed/inline-supplementary-material-1.xlsx?download=true`).

Using the publicly available OTU tables and the publicly available metadata, we estimated the ASV overlap between samples of animal gut microbiota, originating from individual animals of different genera. The

R code that we used to produce these overlap estimates from the publicly available data is provided as Supplemental File 2 and is also available at `https://github.com/FloMazel/Overlap_Animal_Gut_Microbiota`. We proceeded as follow: first, we assigned taxonomy to the ASVs using the RDP classifier and the SILVA v132 database [5] and removed ASVs identified as Chloroplasts or Mitochondria. Second, we merged the ASV tables downloaded from the Qiita website and randomly selected one sample per host species. Third, we rarefied the data to 5000 reads per sample. Fourth, we computed overlap between samples using the presence/absence-based Jaccard metric [6].

We found that, on average, two individuals from different host genera belonging to different classes (e.g. one Mammalian sample and one Avian sample) share 1.26% of their ASVs and two individuals from different host genera belonging to the same class (e.g. two mammalian samples) shared 2.84% of their ASVs (Supplemental File 3).

References

[1] Song SJ, Sanders JG, Delsuc F, Metcalf J, Amato K, Taylor MW, et al. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. MBio. 2020; 11(1). doi:https://doi.org/10.1128/mBio.02901-19.

[2] Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017; 551:457–463. doi:https://doi.org/10.1038/nature24621.

[3] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology. 2006; 72(7):5069–5072. doi:https://doi.org/10.1128/AEM.03006-05.

[4] Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. MSystems. 2017; 2(2). doi:https://doi.org/10.1128/mSystems.00191-16.

[5] Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. Journal of Biotechnology. 2017; 261:169–176. doi:https://doi.org/10.1016/j.jbiotec.2017.06.1198.

[6] Real R, Vargas JM. The probabilistic basis of Jaccard's index of similarity. Systematic Biology. 1996; 45(3):380–385. doi:https://doi.org/10.2307/2413572.