**Title:** Using QIIME2 *Diversity* plug-in for alpha-diversity measurements in microbiome studies

Kaitlyn M. Murphy

A recently published article in *Wildlife Biology* details a study were researchers collected fecal samples from European red deer (*Cervus elaphus*) free-ranging individuals in the Bavarian Forest National Park (with a natural food supply) and those housed in over-wintering enclosures with a supplemented food supply (Menke et al. 2019). 16S rRNA gene-sequencing was conducted to elucidate bacterial taxa, alpha and beta diversities, and abundances of operational taxonomic units (OTUs) from both groups. Importantly, the authors sought to describe the microbiome of the red deer in the park, identify any bacterial differences between deer housed in over-wintering enclosures and those that are free-ranging, and lastly, to document any zoonotic pathogenic bacteria that may reside in the red deer.

Calculating diversity of gut bacteria is critically important for the authors’ goals of this study because of potential variation in richness and evenness values within the subsets of deer. Thus, utilizing two measures of diversity (alpha and beta) ensures sufficient coverage of variation. Alpha-diversity measures the differences within a group of species or taxa; however, beta-diversity measures the differences between a group of species or taxa. In Menke et al., statistical analyses for diversity were conducted in QIIME2 (Quantitative Insights into Microbial Ecology (Caporaso et al. 2010) and called on different ‘plug-ins’ that compute values based on user-specific data. One of these plug-ins, *diversity*, utilizes various functions to calculate different measures of diversity. In particular, the alpha-phylogenetic function measures the genetic component, or phylogenetic component, of taxonomic diversity.

This plug-in can be found online (<https://github.com/qiime2/q2-diversity)> and calls on the alpha-phylogenetic function along with variable options. The required options for the code include an input feature table, phylogenetic tree, the alpha diversity metric (i.e., Faith’s phylogenetic diversity (Faith 1992)), and the name of the output file. This analysis was re-created using the author’s supplied data files (accessed via Dryad Digital Repository) and supplied plug-in code for alpha-diversity.

In reviewing their supplied information, it was impossible to identify which sequences belonged to which subset of deer. A metadata file of which group each sample came from is required to generate further analyses; thus, alpha-diversity measurements were conducted using a subset of sequences from 6 individuals of unknown location. Importantly, while a labyrinth of resources exists on QIIME2 usage, the specific code utilized for their analyses was also not supplied. They reference QIIME2 resources for their subsequent bioinformatics analyses, but do not supply their own code. Thus, reviewers must rely on online QIIME pages for dissecting supplied datasets and the authors’ written directions in the manuscript.

The sequences used in this study were PairedEndFastqManifestPhred33, meaning files contained forward and reverse copies of sequences. A temporary manifest was generated using the sample id’s, directory location, and direction of read (i.e., forward or reverse). The data was trimmed following authors’ instructions in the manuscript using the *dada2* function. Amplicon sequence variants (ASVs) were collapsed and a phylogenetic tree was created using a the ‘masked alignment’ command in QIIME2. The following code was used to generate alpha-phylogenetic measurements of the authors’ data:

*qiime diversity alpha-phylogenetic \*

*--i-table rep\_seq\_feature\_table2.qza \*

*--i-phylogeny rooted-tree.qza \*

*--p-metric faith\_pd \*

*--o-alpha-diversity faith\_pd\_vector.qza*

From the recreated phylogenetic diversity measurements (Fig. 2), I found that most of the values overlapped with what is depicted in the author’s graph (Fig. 1). Most appear to be from Neuhuttenweise or Hochberg locations within the Bavarian forest based on the similarly low values produced. However, S2 and S3 appear to be outliers that might fall within the Hochberg population (natural forest). Again, without knowing which location the sequences were collected from, it is impossible to make comparisons between groups to recreate the authors’ analyses.

Throughout this review process, a lack of code for a variety of manuscripts was identified- namely, those involving microbiome data. Importantly, while data is supplied on online repositories, many discrepancies exist that make analysis impossible. Here, we list additional information that would make secondary analyses on microbiome datasets useful. 1) A metadata file identifying which samples belong to which treatment, population, etc. needs to be supplied. Without this, comparing between groups is impossible because the sequences are not labeled as such. In Menke et al. 2019, the authors’ show a total of 272 forward and reverse sequence files that either correlate with the enclosed or free-ranging individuals. Because they do not supply a metadata file, analyses between the two cannot be completed. 2) Format type of sequences should be supplied in the supporting information. When it is not supplied, this forces reviewers to dissect out the correct format, prolonging downstream analyses. 3) When possible, scripts should be supplied that detail what the authors’ did without relying on manuscript instruction. I think author’s might be hesitant to do so because of the lack of acknowledgements when using supplied scripts. Arguably, the script may have taken the same amount of time (if not longer) as the manuscript itself. Citing references for scripts should also be a more common practice and thus, may promote sharing of bioinformatics resources more widely.

**Source:** <https://github.com/qiime2/q2-diversity>

**Source:** <https://docs.qiime2.org/2020.2/plugins/available/diversity/alpha-phylogenetic/>

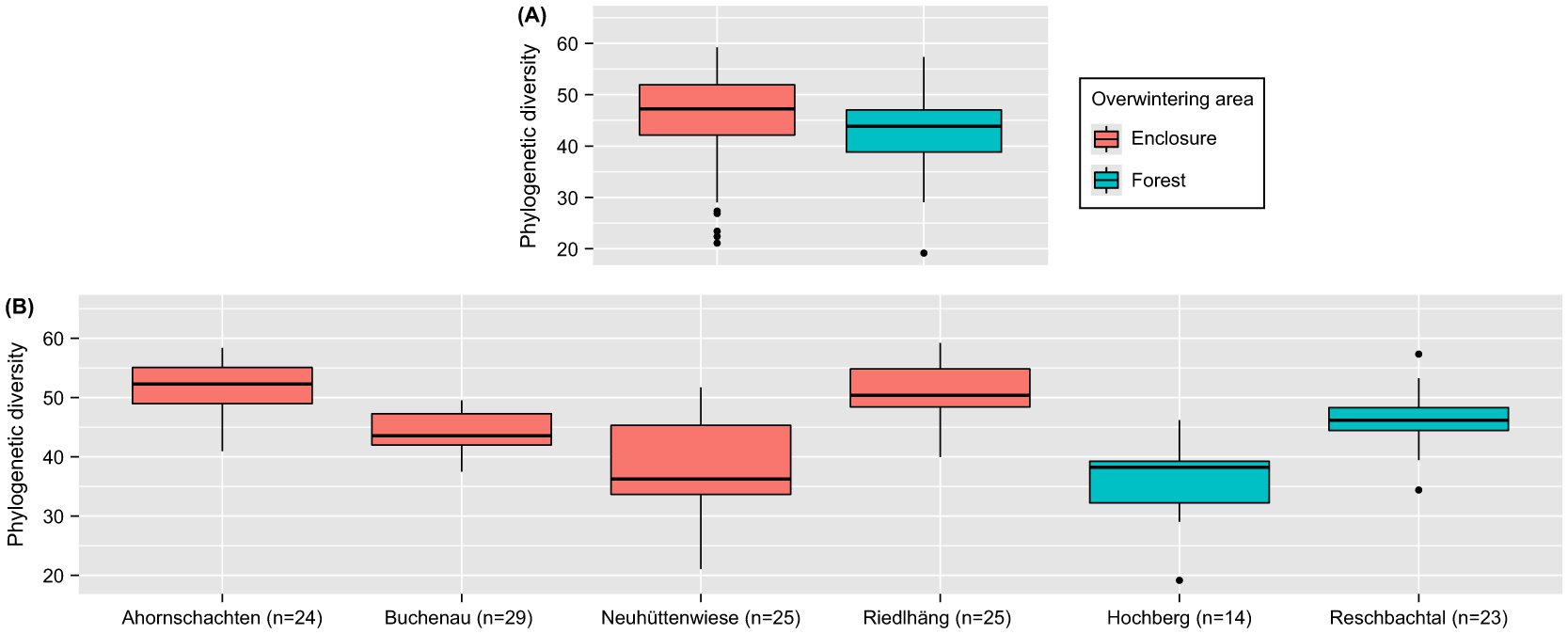
**Data Accessed Via:** <https://datadryad.org/stash/landing/show?big=showme&id=doi%3A10.5061%2Fdryad.7r22vb1>

**References:**

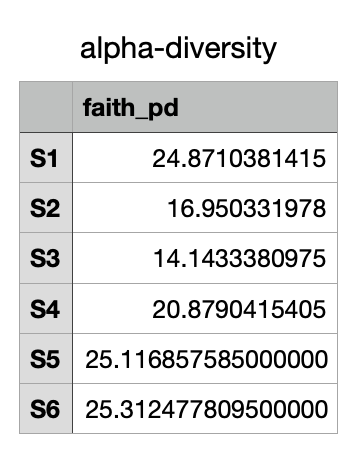
Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7: 335–336. doi:10.1038/nmeth.f.303

Faith DP. Conservation evaluation and phylogenetic diversity. Biological Conservation. 1992;61: 1–10. doi:10.1016/0006-3207(92)91201-3

Menke, S., Heurich, M., Henrich, M., Wilhelm, K., & Sommer, S. (2019). Impact of winter enclosures on the gut bacterial microbiota of red deer in the Bavarian Forest National Park. *Wildlife Biology*, *2019*(1).



**Fig. 1**. Phylogenetic diversity of the red deer gut microbiome between enclosure and forest groups (Menke et al. 2019).



**Fig. 2**. Re-created phylogenetic diversity using accessed data and supplemented code from Menke et al. 2019.