

UC San Diego

UC San Diego Previously Published Works

Title

Determination of Effect Sizes for Power Analysis for Microbiome Studies Using Large Microbiome Databases.

Permalink

<https://escholarship.org/uc/item/07j725b4>

Journal

Genes, 14(6)

ISSN

2073-4425

Authors

Rahman, Gibraan
McDonald, Daniel
Gonzalez, Antonio
[et al.](#)

Publication Date

2023-06-01

DOI

10.3390/genes14061239

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Technical Note

Determination of Effect Sizes for Power Analysis for Microbiome Studies Using Large Microbiome Databases

Gibraan Rahman^{1,2}, Daniel McDonald¹ , Antonio Gonzalez¹, Yoshiki Vázquez-Baeza³, Lingjing Jiang⁴, Climent Casals-Pascual⁵, Daniel Hakim^{1,2}, Amanda Hazel Dilmore^{1,6}, Brent Nowinski⁷ , Shyamal Peddada⁸  and Rob Knight^{1,9,10,*}

¹ Department of Pediatrics, School of Medicine, University of California, San Diego, CA 92093, USA

² Bioinformatics and Systems Biology Program, University of California, San Diego, CA 92093, USA

³ BiomeSense Inc., Chicago, IL 60615, USA

⁴ Janssen Research & Development, Spring House, PA 19002, USA

⁵ Department of Microbiology, Centre de Diagnòstic Biomèdic (CDB), Hospital Clinic, University of Barcelona, 08036 Barcelona, Spain

⁶ Biomedical Sciences Program, University of California San Diego, La Jolla, CA 92093, USA

⁷ Center for Microbiome Innovation, Jacobs School of Engineering, University of California San Diego, La Jolla, CA 92093, USA

⁸ Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences (NIEHS), The National Institute for Health (NIH), Research Triangle Park, Durham, NC 27709, USA

⁹ Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA

¹⁰ Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

* Correspondence: robknight@eng.ucsd.edu

Abstract: Herein, we present a tool called Evident that can be used for deriving effect sizes for a broad spectrum of metadata variables, such as mode of birth, antibiotics, socioeconomics, etc., to provide power calculations for a new study. Evident can be used to mine existing databases of large microbiome studies (such as the American Gut Project, FINRISK, and TEDDY) to analyze the effect sizes for planning future microbiome studies via power analysis. For each metavariable, the Evident software is flexible to compute effect sizes for many commonly used measures of microbiome analyses, including α diversity, β diversity, and log-ratio analysis. In this work, we describe why effect size and power analysis are necessary for computational microbiome analysis and show how Evident can help researchers perform these procedures. Additionally, we describe how Evident is easy for researchers to use and provide an example of efficient analyses using a dataset of thousands of samples and dozens of metadata categories.

Keywords: bioinformatics; microbiome; statistics; effect size



Citation: Rahman, G.; McDonald, D.; Gonzalez, A.; Vázquez-Baeza, Y.; Jiang, L.; Casals-Pascual, C.; Hakim, D.; Dilmore, A.H.; Nowinski, B.; Peddada, S.; et al. Determination of Effect Sizes for Power Analysis for Microbiome Studies Using Large Microbiome Databases. *Genes* **2023**, *14*, 1239. <https://doi.org/10.3390/genes14061239>

Academic Editor: Stefano Lonardi

Received: 1 May 2023

Revised: 24 May 2023

Accepted: 26 May 2023

Published: 9 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Power analysis for a univariate (or multivariate) outcome variable is not new. Numerous statistical packages are available (e.g., SAS) for a variety of experimental designs and outcome variables. For a given level of significance, a common challenge with any power analysis is the understanding of the underlying variability in the data and the value of the parameter of interest in the alternative hypothesis. Once the statistical parameter of interest is identified, researchers often conduct a pilot study to estimate mean differences and standard deviations and use these values, termed effect sizes [1–3], as the basis for conducting power analysis, i.e., sample size calculations for the larger study proposed in their research program. The larger the effect size, the stronger the statistical difference, and the fewer samples are needed for high statistical power.

This type of power analysis is important because of the limited resources available for experimental designs. Ensuring that researchers do not spend more resources than required

to achieve a given statistical power is paramount. The problem is more complicated when it comes to microbiome studies because there are a variety of parameters one can base their designs on. Almost all parameters of interest, such as measures of α or β diversity are (nonlinear) functions of relative abundances of various taxa. Estimation of relative abundances using small pilot studies (i.e., $N < 100$) is not always satisfactory because the observed count data contain a large number of zeros. The preliminary estimates from a pilot study are potentially subject to large bias and uncertainties. Consequently, the determination of the effect size for a given parameter, say α diversity defined by Shannon's entropy, is a difficult task. This article takes the first step towards addressing this challenging problem by making use of the recently created large databases such as the American Gut Project [4], TEDDY [5], and FINRISK [6]. These are very rich databases that continue to grow. They contain microbiome data on several thousands of individuals along with hundreds of commonly measured metadata and thousands of represented taxa. For each variable in the metadata, say, mode of birth, the user-friendly software Evident derives the effect size for a parameter of a researcher's interest, such as Shannon's entropy. Using this parameter, a researcher can then conduct a simulation study to derive power functions for different sample sizes [7].

Since microbiome datasets such as AGP, TEDDY, and FINRISK are very large and contain a large number of metavariables, we expect Evident to be a useful tool for deriving effect sizes for variables of common interest. Importantly, Evident takes user-inputted study data for the generation of results, so researchers can customize their analyses as they see fit. As new databases get constructed, Evident will access those to derive better and more refined effect size estimates that will be useful for planning microbiome studies. Evident is available both as a standalone Python package as well as a QIIME 2 plugin [8]. Currently, effect size analysis and power analysis for microbiome science can be performed using programming languages such as Python and R. However, these approaches are not designed for use with many metavariables. As a result, researchers must write custom code to iterate through the full dataset. With Evident, researchers can seamlessly explore the effect size of community differences in dozens of metadata columns at once and easily perform power analysis. The interactive component of Evident additionally makes this process easy to use and share. This scalability, flexibility, user-friendliness, and integration with existing microbiome software make Evident easier to slot into existing microbiome workflows over existing methods.

2. Materials and Methods

Figure 1a shows an overview of the Evident workflow. As input, Evident takes a sample metadata file and a data file of interest (for example, α diversity). In this cartoon example [9], we show the main Evident workflow is (1) calculating effect size for a metavariable of interest between two or more groups (2) performing parametric power analysis on varying sample sizes, levels of significance, and/or effect sizes (3) plotting the accompanying power curve(s). Both univariate per-sample data (such as α diversity) and multivariate data (as a distance matrix such as β diversity) are supported. For univariate measures, the differences in means among groups are considered. For multivariate measures, the difference in means among within-group pairwise distances is considered. We also note that, at the moment, Evident implements effect size computations of univariable analyses (without explicit handling of confounders) following the approach of existing work [10–13].

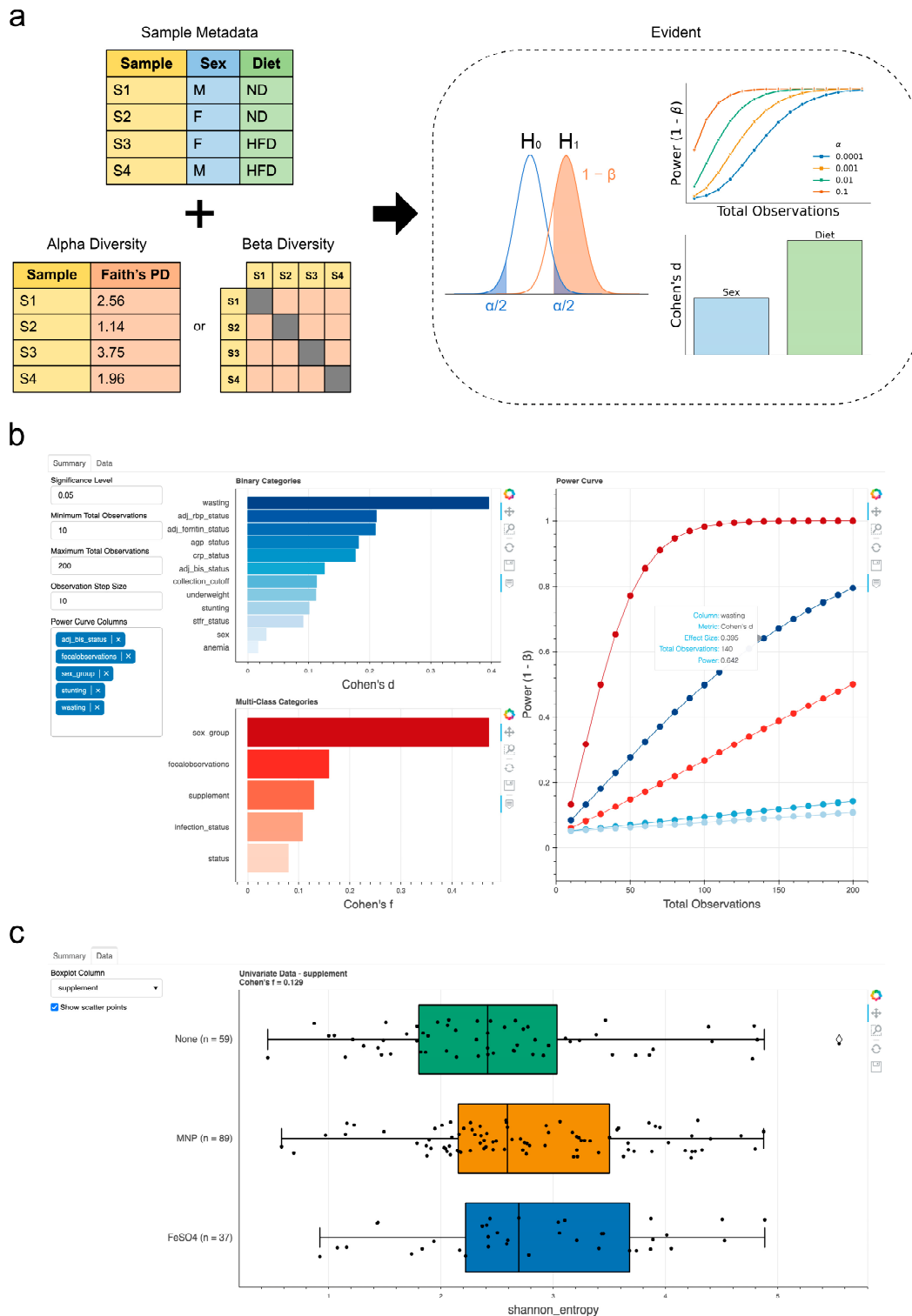


Figure 1. Evident workflow and interactive visualizations. (a) Graphical overview of Evident usage. Sample metadata with categorical groups are used to determine differences among samples. Effect size calculation can be performed and used to generate power curves (in this example using classification status from [7]) at multiple statistical significance levels and sample sizes. (b,c) Screenshots of the interactive webpage for a dynamic exploration of effect sizes and power analysis. Summarized effect sizes of all columns can be used to inform interactive power analysis on multiple groups (b). The underlying grouped data can be visualized with boxplots and, optionally, the raw data as scatter plots (c). The data shown are from McClorry et al. (Qitta study ID: 11402) [9].

Evident supports both binary categories and multi-class categories. For binary categories, Cohen's d is calculated between the two levels. For multi-class categories, Cohen's f is calculated among the levels [14]. Users also have the option of performing pairwise effect size calculations between levels of a multi-class category rather than comparing all groups together. Effect size calculations can be performed on multiple categories at once with simple parallelization by providing the number of CPUs to use. For example, this architecture allows us to decrease the runtime of effect size calculations for 9495 samples comprising 61 categories from over 12 min to 3.5 min using 8 CPUs in parallel.

Evident also provides an interactive component by which users can dynamically explore sample groupings. In Figure 1b,c, we show a screenshot of a web app that users can access with Evident. Metadata categories are pre-sorted by effect size, allowing efficient determination of interesting categories. Power analysis is implemented dynamically—multiple categories can be visualized simultaneously for a specified significance level and number of observations. Researchers can look at the “elbow” of the power curves to determine an optimal number of samples to achieve the desired statistical power for their experiments.

2.1. Statistical Methodology

Let X_1, X_2, \dots, X_l denote l metavariables available in some database. Without loss of generality, in the following, we shall describe the methodology used in Evident for X_1 . For simplicity of exposition, we shall drop the subscript 1 from X_1 . Furthermore, to fix ideas of the methodology and simplicity of exposition, we shall assume X is binary, such as mode of delivery. The outcome variable is denoted by Y , such as Shannon entropy, a measure of α diversity of an infant's gut microbiome. The relative abundance of the j^{th} taxon, $j = 1, 2, \dots, q$, in the k^{th} infant belonging to the $X = i^{\text{th}}$ group, $i = 1, 2, \dots, G$, (e.g., mode of delivery), is denoted by p_{ijk} . For example, $X = 1$ represents babies born vaginally, and $X = 2$ represents babies born by C-Section. We assume that there are q taxa measured on each infant (some may be zeros) and there are N_i infants in the i^{th} the group in the large database. Thus, the Shannon entropy for the k^{th} subject belonging to the $X = i^{\text{th}}$ group is given by $Y_{ik} = -\sum_{j=1}^q p_{ijk} \ln(p_{ijk})$. In this definition, $p_{ijk} \ln(p_{ijk}) \rightarrow 0$, as $p_{ijk} \rightarrow 0$. Since we are working with very large databases, such as AGP, we assume that each N_i is sufficiently large.

The Evident methodology for determining the effect size needed for conducting power analysis and sample size calculations for a future infant gut microbiome study using Shannon entropy to describe microbial diversity is described in the following steps.

Step 1 (Average population diversity): For each value of X , for each subject in the database, using the available microbiome data, compute the desired parameter of interest, for example, the average Shannon entropy for α diversity, $\mu_i = -\left(\frac{1}{N_i}\right) \sum_{k=1}^{N_i} Y_{ik}$, $i = 1, 2$. As noted above, we assume that each N_i is sufficiently large so that μ_i represents the average Shannon entropy, for the i^{th} the population of infants.

Step 2 (Variance of population diversity): Similar to the population mean μ_i , for each $i = 1, 2, \dots, G$ we compute the population variance of α diversity, denoted by $\sigma_i^2 = \left(\frac{1}{N_i}\right) \sum_{k=1}^{N_i} (Y_{ik} - \mu_i)^2$. Again, each n_i is sufficiently large so that σ_i^2 represents the population variance of Shannon entropy for the i^{th} population of infants. Under the simplifying assumption of homoscedasticity (i.e., all populations have same the variance), we average the two empirical variances to obtain the pooled variance, i.e., $\sigma_{pool}^2 = \frac{\sum_{i=1}^G N_i \sigma_i^2}{\sum_{i=1}^G N_i}$. Again, since the sample sizes are large, we regard the pooled variance as the true population variance for all our calculations.

Step 3 (Effect size calculations): Assuming that the outcome variable of interest (e.g., α diversity) is normally distributed, we have the following formulas for effect sizes using

non-central distribution for the test statistic (for $G = 2$) or non-central F distribution (for $G \geq 2$), respectively:

$$d = \frac{\mu_1 - \mu_2}{\sigma_{pool}}$$

$$f = \frac{\sum_{i=1}^G \left(\frac{N_i}{N}\right) (\mu_i - \bar{\mu})^2}{\sigma_{pool}^2}, \quad \bar{\mu} = \sum_{i=1}^G \frac{N_i}{N} \mu_i, \quad N = \sum_{i=1}^G N_i.$$

Although equal variances across groups may be an unreasonable assumption, in the following we make a simplifying assumption that all groups have a common variance of σ_{pool}^2 .

Step 4 (Power and sample size calculations): For a future study, suppose a researcher has a budget for a sample size m_i , for the i^{th} population of infants, $i = 1, 2, \dots, G$, then for a level of significance of α , the power corresponding to the effect size d and sample m_i , can be calculated parametrically, assuming $Y_{ik} \sim^{iid} N(\mu_i, \sigma_{pool})$.

Under the normality assumption, for $G = 2$, Evident calculates power using non-central t-distribution using the effect size parameter d and different choices of samples sizes m_1, m_2 . In the case $G > 2$, it uses non-central F distribution with effect size parameter f and different choices of samples sizes m_1, m_2, \dots, m_G .

2.2. Interactive Exploration of Community Differences

The interactive visualization provided in Evident is created with Bokeh. Given microbiome data and sample metadata, Evident creates a Bokeh app that dynamically calculates effect sizes and power analysis for the chosen parameters. This view also shows the raw data values as boxplots with optional scatter points.

2.3. Analysis of AGP Data

A sample ID list was generated from the original distance matrix used in the AGP study. 100 nucleotide 16S rRNA gene amplicon (16S) data targeting the V4 hypervariable region for these samples were downloaded from the AGP study on Qiita (study ID: 10317) using redbiom [15,16]. Both preparation and sample metadata were also retrieved with redbiom. Due to multiple preparations containing data from some samples, we performed disambiguation by keeping the samples with the highest sequencing depth.

We then processed the feature table and metadata according to the original study. The original workflow used the default parameters in Deblur to remove features with fewer than 10 occurrences in the data [17]. Because Qiita does not perform this filtering by default, we performed this filtering manually. To remove sequences associated with sample bloom, we performed bloom filtering [18]. We then rarefied the feature table to 1250 sequences as in the original analysis.

We processed the sample metadata in accordance with the original study. Because of differences in self-reporting protocols from 2018, metadata categories associated with reported Vioscreen responses as well as those associated with alcohol consumption were removed. The following categories were removed due to mismatches in sample metadata: roommates, allergies, age_cat, bmi_cat, longitude, latitude, elevation, height_cm, collection_time, and center_project_name. Only the top four annotated countries were considered—US, UK, Australia, and Canada. All other countries were ignored. Overall, 61 metadata categories common to both the original data and redbiom data were used for further analysis.

Sequences from the feature table were placed into a 99% Greengenes [19] insertion reference tree using SEPP [20]. We then used unweighted UniFrac to generate a sample-by-sample distance matrix [21]. This distance matrix was used as input to Evident along with the disambiguated, processed sample metadata.

We used effect_size_by_category to calculate the whole-group effect sizes for each column in the metadata and pairwise_effect_size_by_category to calculate the group-pairwise

effect sizes for multi-class categories. For each whole-group effect size, we computed a power analysis for α values of 0.01, 0.05, and 0.1. Power was calculated on total sample size values from 20 to 1500 in increments of 40 samples. Evident analyses were performed in parallel on a high-performance computing environment. Group-wise and pairwise effect size calculations both took under 4 min for 82 metadata categories on 9495 samples using 8 CPUs (we note the AGP paper used $n = 9511$ but operated at 125 nt; we observe a slightly reduced number of samples at 100 nt). We also benchmarked group-wise effect size calculations using only a single CPU as a comparison; this process took 12.4 min, meaning the parallelization decreased runtime by approximately $3.5\times$. Power analysis calculation took 2.7 min for 82 categories using 8 CPUs in parallel.

2.4. Analysis of Study of Latinos Data

We downloaded closed-reference (picked against Greengenes 97%) 16S-V4 fecal data from Qiita (study ID: 11666) using redbiom. We used the `bmi_v2` column to separate samples into two groups: normal (BMI < 25) and obese (BMI > 40). For each sample, we summed the abundance of *Prevotella* spp. and *Bacteroides* spp. adding a pseudocount to both sums. We then calculated the (log) ratio of the *Prevotella* sum and *Bacteroides* sum.

For power analysis, we first established the “true” difference between the obese and normal samples as 1.06 ($d = 0.27$). We used the log-ratio data to determine three levels of effect sizes we wanted to evaluate: 0.5 ($d = 0.13$), 1.0 ($d = 0.25$), and 1.5 ($d = 0.38$). To convert the differences into desired effect sizes, we divided each difference by the pooled standard deviation of the original log ratios. We used Evident to compute the power at each of these effect sizes for a significance threshold of 0.05 for total observations varying between 100 and 1000.

3. Results

As a demonstration of Evident, we reprocessed 9495 samples from the AGP to compare the published effect sizes in McDonald 2018 with those from a new analysis with Evident [4]. We downloaded the same samples from the original paper and reprocessed the data and metadata in the same manner, focusing on within-group UniFrac [22] distances. First, we computed the group-wise effect sizes for all valid metadata categories. The top ten binary categories and multi-class effect sizes are shown in Figure 2a,c, respectively. Using these effect sizes, we performed power analyses for each category at a significance level of 0.05 for a range of sample sizes from 20 to 1500 (Figure 2b,d). We plotted the distribution of the highest effect size binary and multi-class categories as reported by our new analysis in Figure 2e. Finally, we computed the pairwise effect sizes as performed in the original paper to verify that Evident returns the same values. Figure 2f shows that the effect sizes map extremely closely between the published data and the newly reprocessed data. The values of effect size differences in Figure 2g are distributed around 0, indicating that there is very little difference between effect size calculations. This serves as validation that Evident returns the correct effect sizes. We note, however, that the data used in this study is very heterogeneous—coming from multiple countries. It is important to make sure the data used in computing effect sizes are specific to the biological questions of interest. In Supplementary Figure S1, we plot the effect sizes calculated from only US samples and only UK samples. These effect sizes have a weak correlation (Spearman $\rho = 0.54$), suggesting that country is a strong factor for effect sizes between biological groups. As such, researchers may want to perform further pre-processing such as stratification of data by country and computing the individual effect sizes for each population. We believe more work should be done on evaluating these differences in relation to these heterogeneous populations to ensure results are not artificially inflated or deflated.

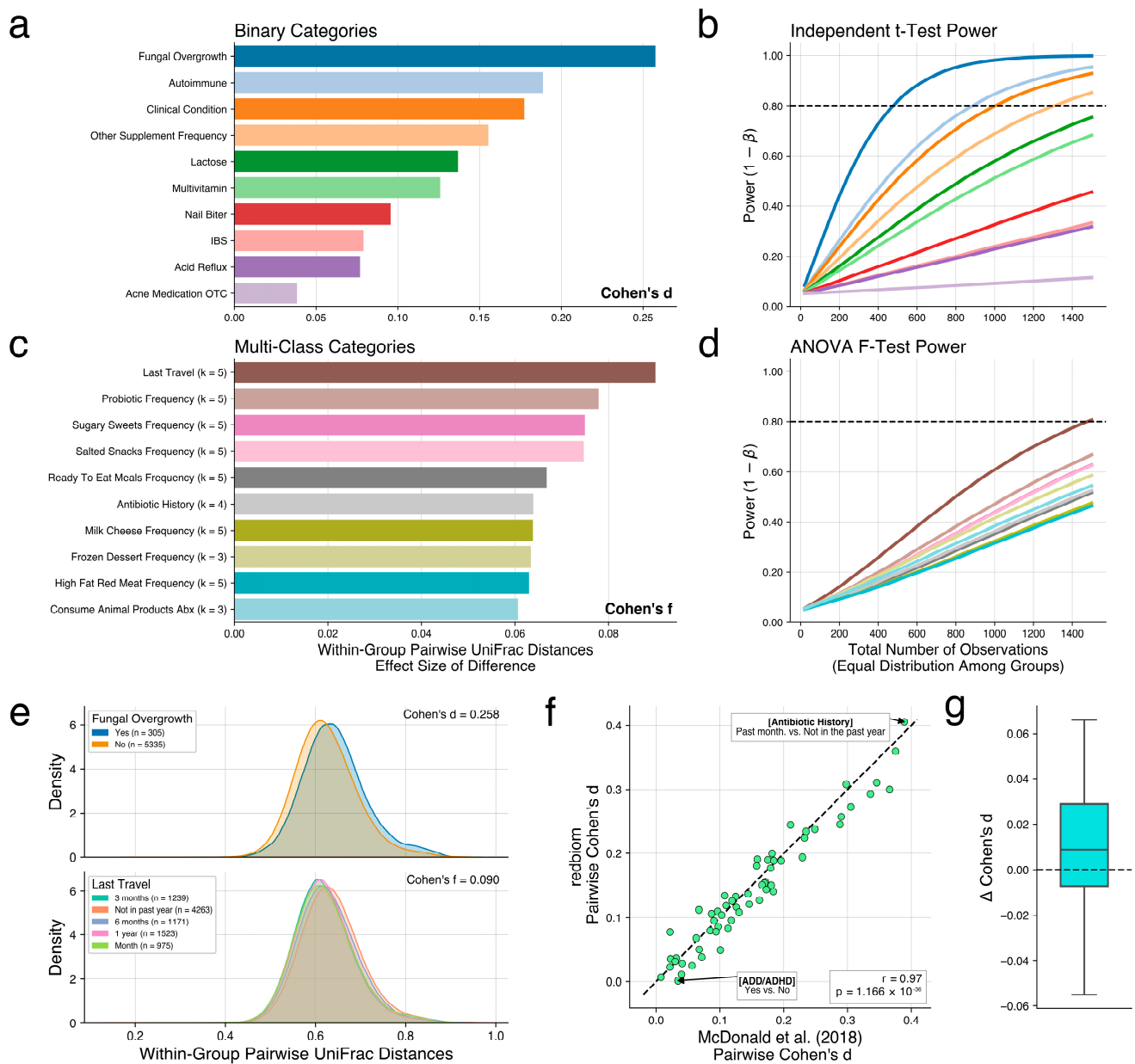


Figure 2. Analysis of American Gut Project data. (a) Top 10 binary categories by group-wise effect size. (b) Two-sample independent t-test power analysis of selected binary category effect sizes for a significance level of 0.05. (c) Top 10 multi-class categories by group-wise effect size. (d) One-way ANOVA F-test power analysis of selected multi-class category effect sizes at a significance level of 0.05. (e) Distributions of within-group pairwise UniFrac distances for highest effect size binary category (top) and multi-class category (bottom). (f) Comparison of pairwise effect sizes between reprocessed data from redbiom and published effect sizes from McDonald et al. Reprocessing results are not identical due to inherent randomness in rarefaction. (g) Boxplot of differences in effect sizes between published and reprocessed effect sizes.

While we focus on diversity measures in this work, Evident is also usable with any other data such as log ratios of microbial abundances. As an example, we use Evident to extend the work of Morton et al. [23] and Fedarko et al. [24] in using log ratios for, e.g., post-hoc differential abundance analysis. We analyzed the commonly reported (log) ratio of *Prevotella* to *Bacteroides* in the Study of Latinos (SoL) cohort [25]. In Supplementary Figure S2, we plot the log ratio differences between subjects with a BMI < 25 and subjects with a

BMI > 40. We also plot a power curve with custom differences in means, showing Evident's flexibility in designing experiments with specific effect sizes in mind.

4. Conclusions

It is important for researchers to keep effect sizes in mind when performing computational microbiome analysis. Calculating and reporting effect sizes make it easier for researchers to determine the magnitude of biological effects on microbial communities. Additionally, these effect sizes can be used to inform power analyses for the efficient allocation of resources for new studies. We designed Evident for researchers to easily mine and process existing datasets for this information. Evident can slot into existing microbiome workflows and process numerous metadata categories efficiently and quickly, allowing its application to a broad range of microbiome research questions.

We note that the choice of study used in Evident should be carefully considered when designing and planning new experiments. For example, an existing study using 16S sequencing may not be completely appropriate when planning shotgun metagenomics experiments, or even experiments that will use a different primer set to target a different region of the 16S rRNA gene, because the different methods may recapture different bacteria with different efficiencies and therefore the effect size of the same per-subject or per-sample variable may differ depending on the methodology. More work should be done to evaluate the differences in downstream analyses on samples between 16S and shotgun metagenomics data. Similarly, culture-based microbiome studies may not follow the same statistical properties as NGS data. Researchers should be mindful of these differences when using Evident. Additionally, researchers should be aware of the limitations of the statistical methodology of Evident. For example, if the assumptions of variance homogeneity are not held, the obtained effect sizes will be inaccurate and the subsequent power analyses can overestimate or underestimate the number of samples required to achieve a given level of statistical power. Similarly, the assumption of equal group sizes in proposed experimental designs from power analysis may be naïve in practice. For rare diseases or phenotypes, it may not be feasible to design an experiment in which all groups have the same number of samples. In these cases, performing simulations with unequal group sizes to determine the necessary sample size to be likely to achieve a statistically significant result may be informative.

We encourage microbiome researchers to incorporate Evident into their workflows for both reporting effect sizes of microbial community differences and planning experimental designs. In the future, we hope to enhance flexibility by including quantitative metadata categories (rather than the current qualitative categories) and unbalanced group sample size power analyses.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14061239/s1>, Figure S1: Comparison of effect sizes between UK and US samples; Figure S2: Log-ratio analysis on Study of Latinos cohort.

Author Contributions: G.R., A.G., D.M. and R.K. conceived the idea for the software and study. G.R., A.G., D.M., Y.V.-B. and L.J. developed the software. G.R., D.M. and A.G. conducted the analysis of the AGP data. B.N. assisted with AGP metadata curation. A.H.D., Y.V.-B., D.M. and D.H. reviewed the software code and provided valuable feedback and bug reports. C.C.-P., L.J. and S.P. contributed to the original code for effect size and power calculation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the Alfred P. Sloan foundation (G-2017-9838), NIH-NIDDK (P01DK078669), NIH-NCI (U24CA248454), and NIH (1DP1AT010885, U19AG063744). Research of S.P. was supported [in part] by funding from NIEHS intramural program ZIAES103389-01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for the demonstration in Figure S1 were downloaded from Qiita (study ID: 11402) [9] at 90 nucleotides using the deblur pipeline. AGP data were downloaded from Qiita (study ID: 10317) using redbiom with context “Deblur_2021.09-Illumina-16S-V4-100nt-50b3a2”. The original pairwise effect sizes, sample metadata, and unweighted UniFrac distance matrix were downloaded from the original McDonald et al. study for comparison. SoL data used in Supplemental Figure S2 were downloaded from Qiita (study ID: 11666) using redbiom with context “Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-90nt-44feac”.

Acknowledgments: We would like to thank the members of the Knight Lab for feedback on the scope and details of Evident. We thank Jamie Morton for valuable discussions about effect size. The authors thank Huang Lin (NIEHS) and Mikyeong Lee (NIEHS) for their valuable comments that improved the presentation of this article.

Conflicts of Interest: The authors declare no conflict of interest.

Code Availability: The latest version of Evident is available at <https://github.com/biocore/evident> under the BSD-3 license. Evident is installable from PyPI both as a standalone Python 3 package and a QIIME 2 plugin. The scripts used to download and analyze AGP data as well as the processed Evident results are available at <https://github.com/knightlab-analyses/evident-analyses>. Analysis of AGP data in this study was performed with Evident version 0.4.0.

References

- Sullivan, G.M.; Feinn, R. Using Effect Size—Or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **2012**, *4*, 279–282. [CrossRef]
- Baguley, T. Standardized or simple effect size: What should be reported? *Br. J. Psychol.* **2009**, *100*, 603–617. [CrossRef]
- Cohen, J. Statistical Power Analysis. *Curr. Dir. Psychol. Sci.* **1992**, *1*, 98–101. [CrossRef]
- McDonald, D.; Hyde, E.; Debelius, J.W.; Morton, J.T.; Gonzalez, A.; Ackermann, G.; Alexander, A. American Gut: An Open Platform for Citizen Science Microbiome Research. *mSystems* **2018**, *3*, e00031-18. [CrossRef]
- TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann. N. Y. Acad. Sci.* **2008**, *1150*, 1–13. [CrossRef]
- Vartiainen, E.; Jousilahti, P.; Alfthan, G.; Sundvall, J.; Pietinen, P.; Puska, P. Cardiovascular risk factor changes in Finland, 1972–1997. *Int. J. Epidemiol.* **2000**, *29*, 49–56. [CrossRef]
- Casals-Pascual, C.; González, A.; Vázquez-Baeza, Y.; Song, S.J.; Jiang, L.; Knight, R. Microbial Diversity in Clinical Microbiome Studies: Sample Size and Statistical Power Considerations. *Gastroenterology* **2020**, *158*, 1524–1528. [CrossRef]
- Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **2019**, *37*, 852–857. [CrossRef]
- McClorry, S.; Zavaleta, N.; Llanos, A.; Casapía, M.; Lönnerdal, B.; Slupsky, C.M. Anemia in infancy is associated with alterations in systemic metabolism and microbial structure and function in a sex-specific manner: An observational study. *Am. J. Clin. Nutr.* **2018**, *108*, 1238–1248. [CrossRef]
- Yang, L.; Chen, J. A comprehensive evaluation of microbial differential abundance analysis methods: Current status and potential solutions. *Microbiome* **2022**, *10*, 130. [CrossRef]
- Dwiyanto, J.; Hussain, M.H.; Reidpath, D.; Ong, K.S.; Qasim, A.; Lee, S.W.H.; Lee, S.M.; Foo, S.C.; Chong, C.W.; Rahman, S. Ethnicity influences the gut microbiota of individuals sharing a geographical location: A cross-sectional study from a middle-income country. *Sci. Rep.* **2021**, *11*, 2618. [CrossRef]
- Park, J.; Kato, K.; Murakami, H.; Hosomi, K.; Tanisawa, K.; Nakagata, T.; Ohno, H.; Konishi, K.; Kawashima, H.; Chen, Y.-A.; et al. Comprehensive analysis of gut microbiota of a healthy population and covariates affecting microbial variation in two large Japanese cohorts. *BMC Microbiol.* **2021**, *21*, 151. [CrossRef]
- Falony, G.; Joossens, M.; Vieira-Silva, S.; Wang, J.; Darzi, Y.; Faust, K.; Kurilshikov, A.; Bonder, M.J.; Valles-Colomer, M.; Vandeputte, D.; et al. Population-level analysis of gut microbiome variation. *Science* **2016**, *352*, 560–564. [CrossRef]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1988; pp. 274–275.
- Gonzalez, A.; Navas-Molina, J.A.; Kosciulek, T.; McDonald, D.; Vázquez-Baeza, Y.; Ackermann, G.; Dereus, J.; Janssen, S.; Swafford, A.D.; Orchanian, S.B.; et al. Qiita: Rapid, web-enabled microbiome meta-analysis. *Nat. Methods* **2018**, *15*, 796–798. [CrossRef]
- McDonald, D.; Kaehler, B.; Gonzalez, A.; DeReus, J.; Ackermann, G.; Marotz, C.; Huttley, G.; Knight, R. redbiom: A Rapid Sample Discovery and Feature Characterization System. *mSystems* **2019**, *4*, e00215-19. [CrossRef]
- Amir, A.; McDonald, D.; Navas-Molina, J.A.; Kopylova, E.; Morton, J.T.; Zech Xu, Z.; Kightley, E.P.; Thompson, L.R.; Hyde, E.R.; Gonzalez, A.; et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* **2017**, *2*, e00191-16. [CrossRef]
- Amir, A.; McDonald, D.; Navas-Molina, J.A.; Debelius, J.; Morton, J.T.; Hyde, E.; Robbins-Pianka, A.; Knight, R. Correcting for Microbial Blooms in Fecal Samples during Room-Temperature Shipping. *mSystems* **2017**, *2*, e00199-16. [CrossRef]

19. McDonald, D.; Price, M.N.; Goodrich, J.; Nawrocki, E.P.; DeSantis, T.Z.; Probst, A.; Andersen, G.L.; Knight, R.; Hugenholtz, P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **2012**, *6*, 610–618. [[CrossRef](#)]
20. Mirarab, S.; Nguyen, N.; Warnow, T. SEPP: SATé-enabled phylogenetic placement. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* **2012**, *2011*, 247–258. [[CrossRef](#)]
21. McDonald, D.; Vázquez-Baeza, Y.; Koslicki, D.; McClelland, J.; Reeve, N.; Xu, Z.; Gonzalez, A.; Knight, R. Striped UniFrac: Enabling microbiome analysis at unprecedented scale. *Nat. Methods* **2018**, *15*, 847–848. [[CrossRef](#)]
22. Lozupone, C.; Knight, R. UniFrac: A New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2005**, *71*, 8228–8235. [[CrossRef](#)]
23. Morton, J.T.; Marotz, C.; Washburne, A.; Silverman, J.; Zaramela, L.S.; Edlund, A.; Zengler, K.; Knight, R. Establishing microbial composition measurement standards with reference frames. *Nat. Commun.* **2019**, *10*, 2719. [[CrossRef](#)]
24. Fedarko, M.W.; Martino, C.; Morton, J.T.; González, A.; Rahman, G.; A Marotz, C.; Minich, J.J.; E Allen, E.; Knight, R. Visualizing omic feature rankings and log-ratios using Qurro. *NAR Genom. Bioinform.* **2020**, *2*, lqaa023. [[CrossRef](#)]
25. Kaplan, R.C.; Wang, Z.; Usyk, M.; Sotres-Alvarez, D.; Daviglius, M.L.; Schneiderman, N.; Talavera, G.A.; Gellman, M.D.; Thyagarajan, B.; Moon, J.-Y.; et al. Gut microbiome composition in the Hispanic Community Health Study/Study of Latinos is shaped by geographic relocation, environmental factors, and obesity. *Genome Biol.* **2019**, *20*, 219. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.