

Article

Impact of Data and Study Characteristics on Microbiome Volatility Estimates

Daniel J. Park ¹  and Anna M. Plantinga ^{2,*} ¹ Roivant Sciences, New York, NY, 10036, USA² Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267, USA

* Correspondence: amp9@williams.edu

Abstract: The human microbiome is a dynamic community of bacteria, viruses, fungi, and other microorganisms. Both the composition of the microbiome (the microbes that are present and their relative abundances) and the temporal variability of the microbiome (the magnitude of changes in their composition across time, called volatility) has been associated with human health. However, the effect of unbalanced sampling intervals and differential read depth on the estimates of microbiome volatility has not been thoroughly assessed. Using four publicly available gut and vaginal microbiome time series, we subsampled the datasets to several sampling intervals and read depths and then compared additive, multiplicative, centered log ratio (CLR)-based, qualitative, and distance-based measures of microbiome volatility between the conditions. We find that longer sampling intervals are associated with larger quantitative measures of change (particularly for common taxa), but not with qualitative measures of change or distance-based volatility quantification. A lower sequencing read depth is associated with smaller multiplicative, CLR-based, and qualitative measures of change (particularly for less common taxa). Strategic subsampling may serve as a useful sensitivity analysis in unbalanced longitudinal studies investigating clinical associations with microbiome volatility.

Keywords: microbiome volatility; longitudinal microbiome; temporal variability; qualitative changes; quantitative changes



Citation: Park, D.J.; Plantinga, A.M. Impact of Data and Study Characteristics on Microbiome Volatility Estimates. *Genes* **2023**, *14*, 218. <https://doi.org/10.3390/genes14010218>

Academic Editor: Yi-Juan Hu

Received: 4 December 2022

Revised: 10 January 2023

Accepted: 12 January 2023

Published: 14 January 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

The human microbiome is a dynamic community of bacteria, viruses, fungi, and other microorganisms that has been associated with a wide range of human diseases, including irritable bowel disease, graft-versus-host disease, bacterial vaginosis, and Alzheimer's [1–4]. With improvements in sequencing technologies and reductions in cost, longitudinal microbiome studies have become increasingly common in recent years. Longitudinal studies of microbiome composition have been used to investigate temporal changes in healthy individuals [5–8] and to understand the association of changes in microbiome composition with clinical changes during disease progression or treatment [9–13]. In addition to clinical associations with the abundance of specific taxa, microbiome volatility, which we define as the degree of compositional change over time [14], may have a separate and potentially bidirectional association with health [10,15].

Microbiome volatility does not have a single, formal mathematical definition. There are two major approaches to quantifying and comparing microbiome changes over time: distance-based and taxon-level approaches. In distance-based approaches, rather than defining an explicit volatility metric, investigators compute ecological distance metrics such as Bray–Curtis or UniFrac distances between samples [16], then compare the distributions of intraindividual and interindividual distances [14,15,17–19]. For example, Bastiaansen et al. [14] defined volatility as the Aitchison distance travelled over the course of a study and concluded that stressed mice had higher volatility than the controls; similarly, Halfvarson et al. [15] computed weighted and unweighted UniFrac distances between

consecutive samples, set a boundary they called the “healthy plane” to represent the variability within healthy individuals, and found that participants with IBD had “considerable volatility away from the healthy plane”. Distance metrics may also be computed directly based on qualitative or quantitative taxon-level measures of change [20]. Formal taxon-level approaches to characterizing microbiome volatility are less common and typically require a long, dense microbiome time series. For example, measures related to microbiome volatility can be extracted from temporal models such as dynamic linear models [21] and the linear mixed model based approach called MTV-LMM [22]. The intraclass correlation coefficient has also been used to quantify stability [7,23]. In some cases, investigators combine these two conceptual approaches by associating the abundance of specific taxa with intraindividual variability [17].

However, there are several features of longitudinal microbiome studies that may pose challenges to volatility quantification and comparison. First, similar to most longitudinal studies, even microbiome studies that are designed to collect samples at balanced time points are afflicted by missing, mistimed, or QC-failed samples, so the sampling interval is rarely consistent for all intraindividual pairs of samples. In addition, sampling intervals for gut microbiome profiling are often tied to a biological mechanism such as a bowel movement, so timing cannot be precisely controlled by investigators. Second, the perennial difficulty posed by differences in sequencing read depth comes to the fore for comparisons of taxon presence across time points. The absence of rare taxa in a sequencing-based microbiome profile may represent true biological absence or insufficient sampling depth to detect the taxon. Hence, variability in the total read count across repeated samples may result in higher qualitative volatility estimates than would be supported based on a perfect knowledge of the microbial community. Third, microbiome data are compositional, so changes in a taxon’s relative abundance may be driven by changes in that taxon’s absolute abundance, but they may also be induced indirectly by changes in the abundance of other taxa.

Using four publicly available microbiome time series datasets, we investigated the potential impact of sampling interval length and sequencing read depth on several measures of change in microbiome composition. We used temporal subsampling and several rarefaction approaches to systematically vary the sampling interval and read depth, then computed additive, multiplicative, centered log ratio (CLR)-based, and qualitative (presence/absence) changes in the taxon abundance between time points along with the typical intraindividual distance-based analysis. This exploration reveals patterns in the measures of microbiome volatility across the study and sample characteristics, with implications for the design and analysis of microbiome volatility studies.

2. Materials and Methods

2.1. Datasets

The four microbiome time series datasets selected for use in this study varied in their sampling interval, overall study time frame, and body site. We focused on internal body sites (as opposed to the skin, for example, which directly interacts with the external environment) with different microbiome characteristics: the healthy gut microbiome tends to be diverse, whereas the healthy vaginal microbiome tends to have one or a few heavily dominant species. Selected study characteristics are summarized in Table 1 and visualized in Supplemental Figure S1.

Caporaso et al. (2011) generated the densest and longest-running time series of the human gut microbiome to date, called the Moving Pictures study [5]. They sampled two healthy individuals almost daily for 6 months (female subject F4) or 15 months (male subject M3), sequencing the V4 region of the 16S rRNA gene to generate microbiome profiles. We used genus-level taxonomic profiles.

In the Student Microbiome Project (SMP), Flores et al. (2014) sampled 85 college-aged adults at four body sites (gut, forehead, palm, and tongue) weekly for 3 months, sequencing the V4 region of the 16S rRNA gene and clustering sequences into OTUs at 97%

similarity [24]. We used only gut microbiome samples. Demographic, lifestyle, medication, and health status data were also collected weekly; subjects who took antibiotics during the study period were excluded from this analysis. Neither subjects nor samples were excluded on the basis of sickness during the study period (37 out of 58 subjects) or menstruation status. There were 10509 unique OTUs across all the samples, which we aggregated at the genus level for computational tractability.

Gajer et al. (2012) sampled the vaginal microbiome of 32 healthy women twice weekly for 16 weeks, sequencing the V1-V2 region of the 16S rRNA gene [25]. Participants reported menstrual bleeding, sexual activity, medications, contraceptives, and other characteristics in daily diaries. Taxon counts were computed based on the original study's reported taxon proportions and total read counts, rounded to the nearest whole number. Taxonomic assignments were at the species level for the *Lactobacillus* species and genus level otherwise.

Ravel et al. (2013) characterized the vaginal microbiome daily for 10 weeks in 4 women without bacterial vaginosis (BV), 6 women with asymptomatic BV (ABV), and 15 women with symptomatic BV (SBV), sequencing the V1-V3 regions of the 16S rRNA gene and generating species-level taxonomic assignments [26]. For women with episodes of ABV or SBV, we included only samples prior to the BV episode. Notably, women later diagnosed with ABV or SBV typically had *Lactobacillus*-depleted vaginal microbiomes at earlier time points, despite the lack of an active BV diagnosis based on their Nugent score. We excluded women with fewer than 20 time points prior to their first active BV diagnosis.

Table 1. Characteristics of studies included in this investigation. Sample sizes and time points after all necessary exclusions.

	Caporaso et al., 2011 [5]	Flores et al., 2014 [24]	Gajer et al., 2012 [25]	Ravel et al., 2013 [26]
Basic Study Information				
Study name	Moving Pictures	SMP	-	-
Body site	Gut	Gut	Vagina	Vagina
Number of subjects	2	58	32	6
Percent female	50%	63.7%	100%	100%
Percent white	-	75.9%	40.6%	16.7%
Age (years): Mean (SD)	-	24.1 (6.4)	37.1 (8.1)	27.2 (6.3)
Age (years): Range	32–33	18–55	22–53	21–38
Sampling Frequency and Study Duration				
Number of time points	131–336	7–10	25–33	23–38
Sampling interval	Daily	Weekly	Twice-weekly	Daily
Study duration	6–15 months	3 months	16 weeks	10 weeks
Summaries of Taxa and Reads				
Read count: Median	36,114	43,282	2403	5195
Read count: Range	15,355–60,847	11,393–188,192	556–6619	145–15,972
Number of unique taxa	3962	632	331	122
Taxon analysis level	Genus	Genus	Genus/Species	Species

2.2. Measures of Volatility

We considered five approaches to quantifying changes in the microbiome between pairs of time points. The four taxon-level measures of change were additive, multiplicative, CLR-based, and qualitative measures, as summarized in Table 2. We also considered intraindividual global dissimilarities.

Since the logarithm of zero is undefined, the CLR transformation requires replacing zeros with small non-zero values. There are a variety of methods for zero replacement [27–30];

we followed the common approach of adding a pseudocount of 1 to all counts [31]. Notably, d^c is related to d^m through a ratio of geometric means:

$$\begin{aligned}
 d_{ijt_{k-1}t_k}^c &= \log\left(\frac{\tilde{p}_{ijt_k}}{\text{GM}(\tilde{p}_{it_k})}\right) - \log\left(\frac{\tilde{p}_{ijt_{k-1}}}{\text{GM}(\tilde{p}_{it_{k-1}})}\right) \\
 &= \log\left(\frac{\tilde{p}_{ijt_k}}{\tilde{p}_{ijt_{k-1}}} \bigg/ \frac{\text{GM}(\tilde{p}_{it_k})}{\text{GM}(\tilde{p}_{it_{k-1}})}\right) \\
 &= \log\left(d_{ijt_{k-1}t_k}^m \times \frac{\text{GM}(\tilde{p}_{it_k})}{\text{GM}(\tilde{p}_{it_{k-1}})}\right),
 \end{aligned}
 \tag{1}$$

where \tilde{p}_{ijt} is the relative abundance of the taxon j for subject i at time t computed after pseudocount addition and $\text{GM}()$ is the geometric mean. For taxa absent at both time points in a rarefied dataset,

$$\tilde{p}_{ijt_{k-1}} = \tilde{p}_{ijt_k} = \frac{1}{\text{Rarefied Read Count} + \text{Num. Taxa}}.
 \tag{2}$$

In this case, d^c simplifies the log of the ratio of geometric means, so there is often a nonzero CLR-based difference even when the taxon is not detected at either time point.

Table 2. Taxon-level measures of change. For subject i and taxon j at consecutive time points, t_{k-1} and t_k , $d_{ijt_{k-1}t_k}$ indicates the measure of change in taxon abundance between the two-time points. Relative abundance is indicated by p_{ijt} and the centered log ratio (CLR) transformation is defined as $\text{CLR}(\tilde{p}_{ijt_k}) = \log(\tilde{p}_{ijt_k} / \text{GM}(\tilde{p}_{it_k}))$ where \tilde{p} represent proportions computed after pseudo count addition, and $\text{GM}()$ is the geometric mean.

	Definition	Requirements	Possible Values
Additive	$d_{ijt_{k-1}t_k}^a = p_{ijt_k} - p_{ijt_{k-1}}$	-	$[-1, 1]$
Multiplicative	$d_{ijt_{k-1}t_k}^m = \log(p_{ijt_k} / p_{ijt_{k-1}})$	$p_{ijt_k} > 0, p_{ijt_{k-1}} > 0$	$(-\infty, \infty)$
CLR-Based	$d_{ijt_{k-1}t_k}^c = \text{CLR}(\tilde{p}_{ijt_k}) - \text{CLR}(\tilde{p}_{ijt_{k-1}})$	\tilde{p} computed after zero-replacement	$(-\infty, \infty)$
Qualitative	$d_{ijt_{k-1}t_k}^q = I(p_{ijt_k} > 0) - I(p_{ijt_{k-1}} > 0)$	-	-1 (present → absent), 0, 1 (absent → present)

Intraindividual dissimilarities were calculated using existing ecological distance metrics. For the Moving Pictures dataset, which includes a phylogenetic tree, we considered unweighted UniFrac, generalized UniFrac with $\alpha = 0.5$, weighted UniFrac, and Bray–Curtis dissimilarities. The UniFrac dissimilarities account for phylogenetic relationships among taxa, whereas Bray–Curtis only considers taxon abundance [16]. Unweighted UniFrac only uses presence/absence data; weighted UniFrac only uses abundance data; and generalized UniFrac is an intermediate version [32]. For the other three studies, because phylogenetic trees were unavailable, we only calculated the Bray–Curtis dissimilarity.

2.3. Sampling Interval and Sampling Depth Investigations

We considered four sampling intervals: 1 day, 3 days, 7 ± 1 days, and 28 ± 4 days, with the acceptable time lag between the samples designed to keep the observed sampling intervals to within 15% longer or shorter than the desired sampling interval. Once the pairs of samples at the desired sampling interval were identified, all samples were rarefied to the minimum read count in that study, the four taxon-level measures of change (d^a, d^m, d^c, d^q) were calculated for each taxon in each pair of samples, and the global dissimilarity was calculated between the sample pairs. Because the minimum read count was very small for the Ravel study, we treated 500 reads as the “minimum read count” for rarefaction purposes (leaving 5 samples unrarefied). The summary statistics included the standard deviation of d^a, d^m , and d^c and the proportion of time pairs for which $d^q \neq 0$ were used to

compare the magnitude of taxon-level changes based on sampling interval and average taxon relative abundance.

We also considered four rarefaction-based approaches to explore the impact of differential sampling depth. Specifically, we: (1) kept all of the original reads (no rarefaction), (2) adopted the conventional approach of rarefying to the minimum read count across all samples, (3) rarefied to 80% of the minimum read count, and (4) rarefied to 60% of the minimum read count. Using these four versions of each dataset with weekly sampling (7 ± 1 day sampling interval), we repeated the summaries described above. As before, 100% of the minimum read count was considered 500 reads for the Ravel study.

Finally, to assess whether samples that originally had higher read counts had different estimated volatility even after rarefaction, we fixed the sampling interval for each study to a seven-day interval, rarefied to the minimum total read count, and calculated all measures of volatility. We averaged the two original read counts for each time point pair and explored associations of the taxon-level measures of change with the pairwise average read count.

3. Results

3.1. Sampling Interval Investigations

In the absence of systematic microbiome transitions to a new stable state over the course of a study, each measure of quantitative change (additive, multiplicative, CLR-based) should be centered around zero. For the most part, this is what Supplemental Figure S2 shows, although the CLR-based changes for abundant taxa at long sampling intervals (time lags) had a positive mean. Therefore, because comparing the centre of the distribution of log fold changes is not highly informative, we instead compared distributions based on the spread, where larger standard deviations (SDs) indicate larger increases and decreases in relative abundance.

The SDs for additive changes in relative abundance are shown in the top row of Figure 1 and in Supplemental Table S1. In all four studies, the overall SD was larger for longer time lags, and more abundant taxa had much larger SDs than rarer taxa. Taxa with relative abundances greater than 0.001 followed the overall association of higher SDs for longer sampling intervals. However, for rare taxa, the opposite was true: longer sampling intervals had smaller SDs.

For multiplicative changes, shown in Figure 1 (middle row) and Supplemental Table S2, the results were similar to additive changes. For rare taxa, smaller SDs were observed at longer time lags, whereas for common taxa, larger SDs were observed at longer time lags. The multiplicative SDs were shrunk less strongly towards zero than the additive SDs due to rare taxa that are absent at most pairs of time points, which have additive changes of zero but undefined multiplicative changes.

The results for CLR-based changes (bottom row of Figure 1; Supplemental Table S3) revealed consistent increases in SDs both with increasing taxon abundance and with increasing time lag within each relative abundance category. The two factors together had a synergistic effect: the ratio between the SDs of CLR-based changes for larger versus smaller time lags was greater for more abundant taxa.

Sampling interval was relatively unassociated with the probability of qualitative changes in taxon abundance, as seen in Figure 2. Qualitative changes were most common at intermediate taxon abundances, which is consistent with biological expectations: extremely rare taxa will almost never appear, and extremely common taxa will be present at nearly all time points, so qualitative changes are unlikely at both extremes. In the Moving Pictures dataset, the probability of qualitative change was noticeably higher for a 28-day lag than for the three shorter lag times, and similarly, for the Ravel dataset, the probability of qualitative change monotonically increased with increasing sampling interval. Although the SMP and Gajer datasets descriptively followed the same pattern, the differences are smaller.

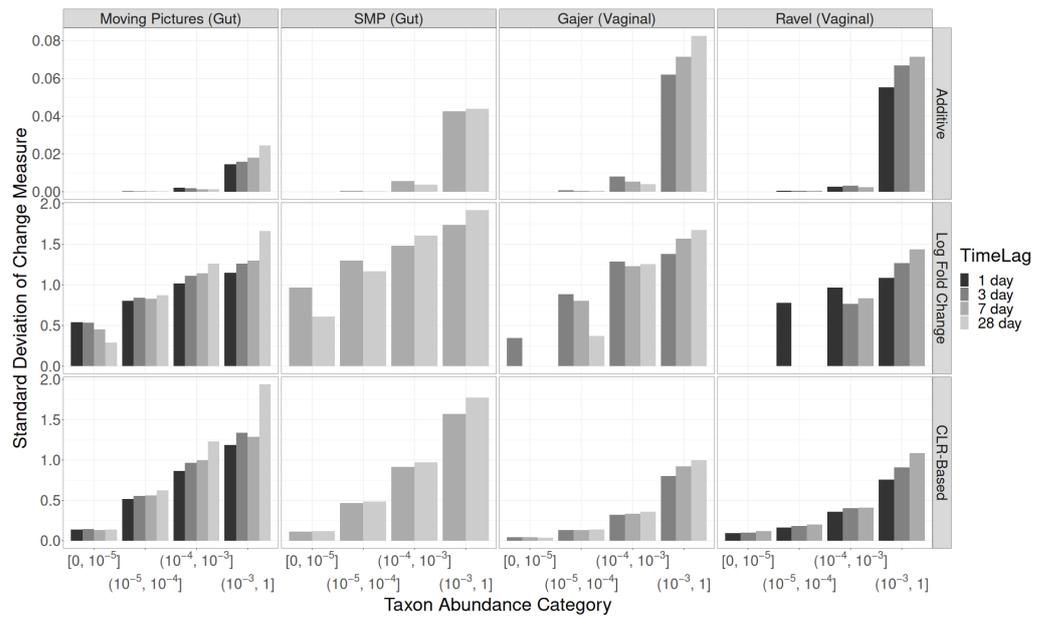


Figure 1. Standard deviations of quantitative taxon-level measures of change by time lag, taxon abundance category, and study. Top row: additive changes. Middle row: multiplicative changes (log fold changes). Bottom row: centered log ratio based changes.

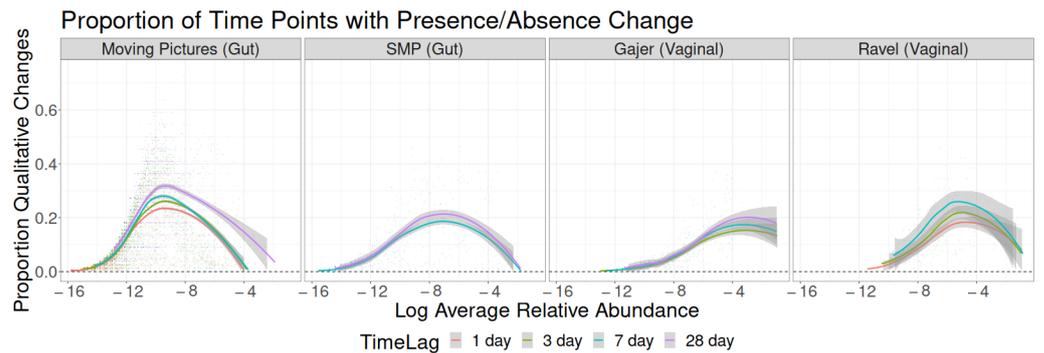


Figure 2. Proportion of time point pairs for which a taxon’s binary presence in the sample changes as a function of study, time lag, and the log of the taxon’s average relative abundance across all samples.

Finally, for intraindividual global dissimilarities, there was a mild increase in the median Bray–Curtis dissimilarity between time point pairs as time lag increased for each of the four studies (Figure 3). The differences were most noticeable for the Moving Pictures dataset, particularly on day 28 (which was also the scenario with the highest rate of qualitative change). Both observations may result in part from the higher diversity of the Moving Pictures dataset, which has far more unique taxa than the other three datasets. To confirm that similar results hold in other common dissimilarity metrics, Supplemental Figure S3 compares four measures of global dissimilarity by time lag in the Moving Pictures dataset. Similar patterns across time lags were seen for each dissimilarity: the distribution of the dissimilarity value for lags of 1 and 3 days were similar, the distribution at 7 days had slightly fewer high outliers, and the distribution at a 28-day lag had higher variability (indicated by a larger interquartile range and heavier right tail).

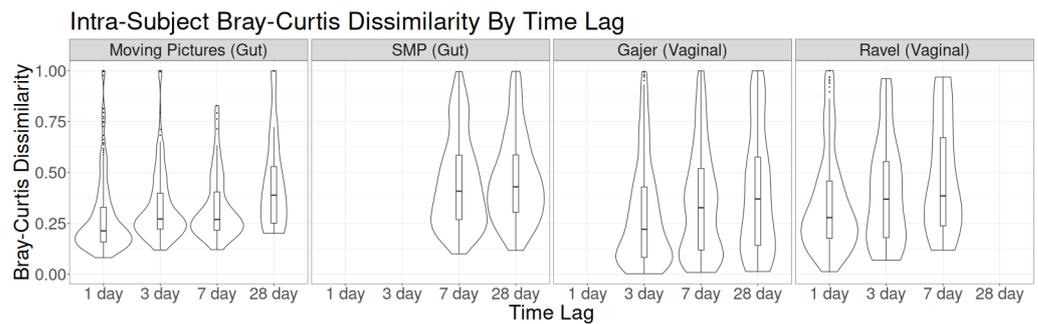


Figure 3. Distribution of intraindividual Bray–Curtis dissimilarity across time lags for each of the four studies.

3.2. Read Depth Investigations

Another key question in longitudinal microbiome studies is how much of an effect differences in read depth has on measured qualitative and quantitative change in microbiome composition and whether that effect persists even after rarefying the data to a common total read count. As in the previous section, the centre of each distribution of changes (with different rarefactions, at a fixed 7-day time lag) is zero, so we again compare SDs.

The SDs of **additive** changes in taxon relative abundance were similar regardless of taxon abundance category, study, and rarefaction procedure (top row of Figure 4; Supplemental Table S4). For **CLR-based** changes (bottom row of Figure 4; Supplemental Table S6), the SD decreased monotonically with increasing strength of rarefaction in each study and across all taxon abundance categories. As expected given the high variability in read counts within each study, the difference between no rarefaction and standard (100%) rarefaction is larger than differences between rarefaction levels. **Multiplicative** changes (middle row of Figure 4; Supplemental Table S5) matched CLR-based changes almost exactly, with the exception of rare taxa in the vaginal datasets (which often disappeared from the dataset entirely after rarefaction).

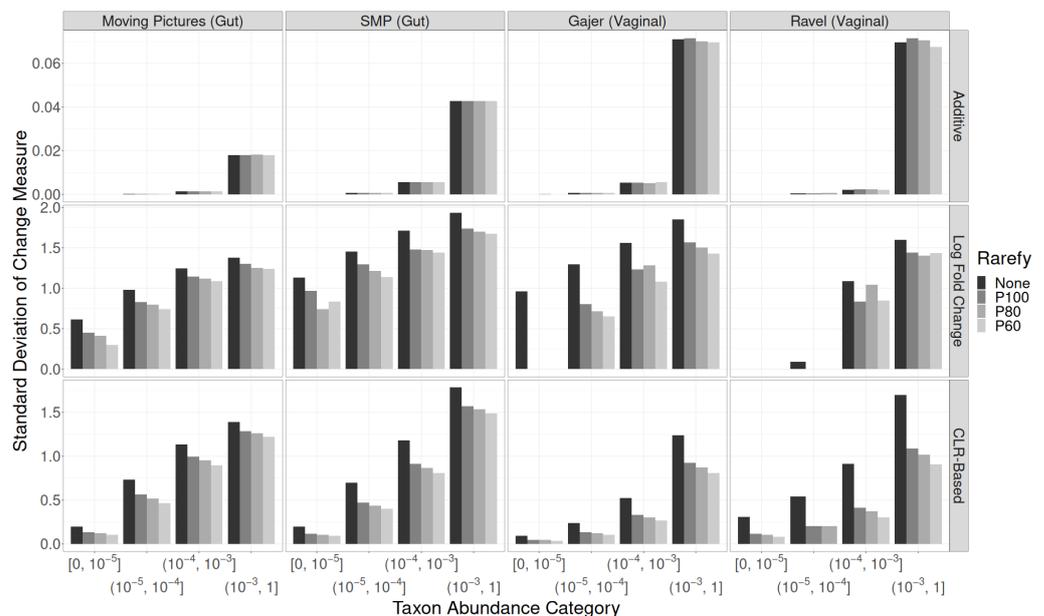


Figure 4. Standard deviations of quantitative taxon-level measures of change by rarefaction procedure, taxon abundance category, and study. Top row: additive changes. Middle row: multiplicative changes (log fold changes). Bottom row: centered log ratio based changes.

Differences in the frequency of qualitative change at each relative abundance occurred only between the unrarefied and rarefied versions of each dataset; the level of rarefaction

did not make a notable difference (Figure 5). In the Moving Pictures, SMP, and Gajer studies, the shape of all four curves was similar, but the unrarefied curve was shifted to the left relative to the rarefied curves so that the maximum probability of qualitative change occurred at a lower abundance. The Ravel study showed a similar pattern but with larger standard errors around the LOESS curve. Because subsampling investigations focus only on sampling zeros (biologically present taxa that have zero counts due to undersampling), not structural zeros (zero counts due to biological absence), these results confirmed that more frequent intermittent sampling zeros, as seen with reduced read counts, were associated with higher estimated qualitative volatility.

The distribution of dissimilarities was nearly identical regardless of rarefaction choices, so mild shifts in the estimates of taxon abundance did not have much impact on the distance-based measures of intraindividual change over time (Figure 6 and Supplemental Figure S4). The only noticeable differences in dissimilarity distributions occurred when unweighted UniFrac is used. In that case, the intraindividual dissimilarity tended to be very slightly higher with lower read counts (increasing rarefaction). Because read depth and rarefaction most strongly affect the presence of rare taxa, not the abundance of common taxa, unweighted UniFrac is by nature the most sensitive to read depth effects due to its exclusive use of presence/absence data.

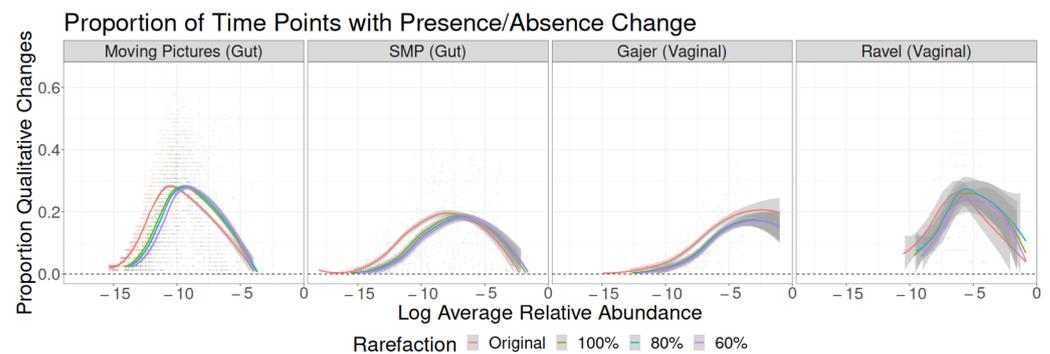


Figure 5. Proportion of time point pairs for which a taxon’s binary presence in the sample changes as a function of study, rarefaction approach, and the log of the taxon’s average relative abundance across all samples.

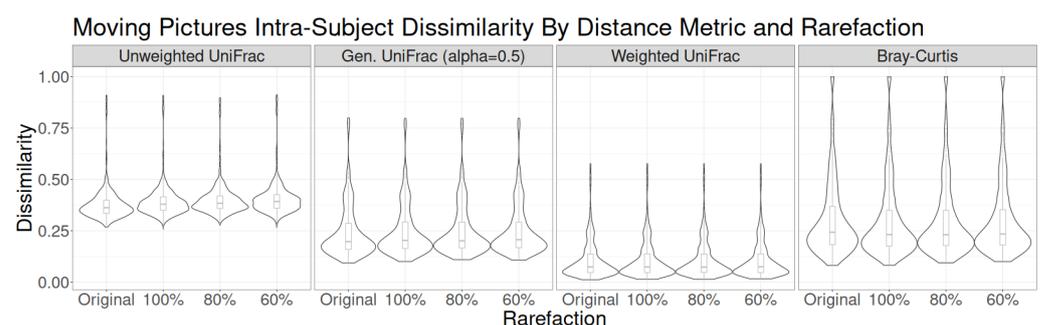


Figure 6. Intraindividual dissimilarity quantified using four metrics (unweighted UniFrac, generalized UniFrac, weighted UniFrac, and Bray–Curtis dissimilarity) for each rarefaction approach in the Moving Pictures study.

Residual Effects of Read Depth after Rarefaction

Whether rarefaction as an approach to correct for differences in sequencing read depth is appropriate, inefficient but valid, or invalid is currently in contention (see, e.g., [33–36]). Some recent investigations have shown that there may be residual effects of differences in read depth even after rarefaction, particularly in analyses strongly affected by the presence of rare taxa, such as alpha diversity analyses [37]. Others show that rarefaction is inefficient but yields proper expected values for the desired quantities [34]. Because rarefaction is still

commonly used in practice, we investigated whether there were residual effects of samples' original read depths on measures of taxon-level change.

For additive changes, the top row of Figure 7 shows that in both gut studies (Moving Pictures, SMP), the SDs for taxa with relative abundance in (0.001, 1] were largest for the quartile of sample-pairs with the highest original read counts; however, in the vaginal studies (Gajer, Ravel), the samples with lower original read counts had higher SDs. Multiplicative changes also tended to have larger SDs for samples with higher original read counts in Moving Pictures, SMP, and Gajer (but not Ravel), for all but the rarest taxon abundance categories (Figure 7, middle row).

For CLR-based changes between consecutive non-zero time points, patterns differed between all four studies (Figure 7, bottom row). The Moving Pictures SDs tended to increase with increasing read count quartile. The SMP SDs tended to be highest for the samples with the lowest and highest quartile of original read counts. Gajer showed no association between SD and the original read count at all, and for Ravel, the samples in the lowest quartile of the original read count had the highest SDs.

Finally, for qualitative changes (Figure 8), the overall takeaways were similar to CLR-based findings. In all studies except Ravel, the samples in the highest read count quartile had the highest qualitative volatility; these differences were more noticeable in the Moving Pictures study than in SMP or Gajer. By contrast, in Ravel, the samples in the two lower read count quartiles tended to have higher qualitative volatility.

Because these potential residual effects were not consistent across the four studies, it is possible that some of the differences occurred by chance based on a particular subsampling of the data. Further investigation in additional time series studies is needed to clarify the extent to which residual read depth effects are a concern in volatility studies and perhaps suggest a better approach than rarefaction to handle them.

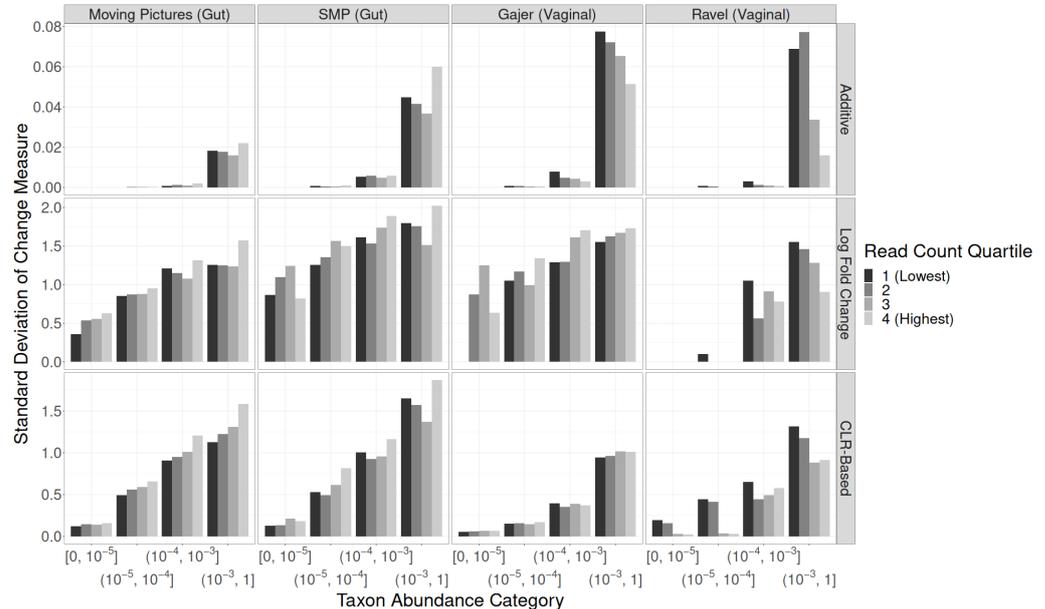


Figure 7. Standard deviations of quantitative taxon-level measures of change by original read count quartile, taxon abundance category, and study. Top row: additive changes. Middle row: multiplicative changes (log fold changes). Bottom row: centered log ratio based changes.

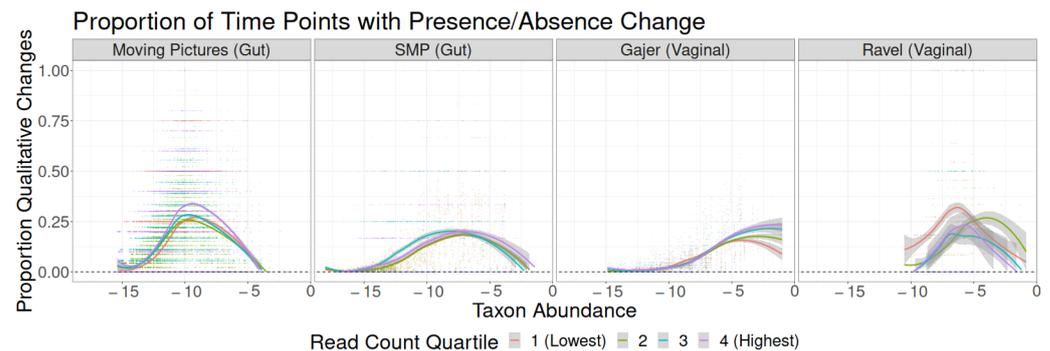


Figure 8. Proportion of time point pairs for which a taxon’s binary presence in the sample changes, as a function of study, original read count quartile, and taxon abundance quartile.

4. Discussion

Through systematic subsampling of four microbiome time series, we have explored the effect of sampling interval on measures of microbiome volatility. Quantitative measures of change, including additive differences in relative abundance, log fold changes in relative abundance, and additive differences in CLR-transformed abundance, tend to be larger at longer sampling intervals, particularly for common taxa. However, the increase in volatility is not constant across relative abundances, so simply using a multiplier such as the inverse time interval is unlikely to fully account for sampling interval-related differences. Sampling interval is generally unassociated with the probability of qualitative change, although in the Moving Pictures study, the proportion of qualitative change is greater at the longest time lags. The probability of qualitative change is largest for intermediate-abundance taxa in every study. Finally, the differences between the distributions of intraindividual dissimilarities across sampling intervals are minor in every study except Moving Pictures, for which dissimilarities are greatest at the 28-day time lag. Comparisons across time intervals are similar for the four distance metrics considered.

We have also explored the effect of read depth on each measure of change in microbiome composition through the rarefaction of each microbiome time series to different read depths with fixed 7-day sampling intervals. In general, additive measures of change are similar regardless of the read depth, whereas multiplicative and CLR-based measures of change are smaller with lower read depths. The distribution of the proportion of qualitative changes shifts to higher relative abundances with lower read depths. There are minimal differences in the distributions of distance-based measures of volatility with rarefaction.

Associations between the measures of change calculated on rarefied data and the original read count of the two samples are inconsistent across the four studies, so more evidence would be needed to conclude that there is a residual effect of sequencing read depth after rarefaction and to clarify what that effect is. However, ongoing investigations into this question are worthwhile, as results may vary depending on the distribution of taxon abundances, the measures being calculated post-rarefaction and their dependence on the presence of rare taxa, and other study factors.

Across all investigations, the distribution of differences in CLR-transformed relative abundances is relatively similar to the distribution of log fold-changes in relative abundances. Given the relationship between these two measures (Equation (1)), this behaviour is expected if the geometric mean taxon abundance is not extremely variable between the samples. Since using the CLR transformation mitigates the effects of compositionality in taxon-level analyses [38] and the requisite zero-adjustment allows absent taxa to contribute to multiplicative measures of volatility while maintaining results consistent with the raw log ratio among pairs of non-zero taxon abundances, the CLR-based approach is an attractive measure of quantitative microbiome volatility.

Taken as a whole, our results suggest that the effects of sequencing read depth and sampling interval are consistent across studies and body sites after accounting for taxon relative abundances. Sequencing read depth, consistent with general intuition, matters most

for rare taxa and qualitative changes, and rarefaction helps avoid systematic differences in volatility estimation based on read depth. This comes at the cost of efficiency due to the reduction in the available data; more sophisticated methods, including some modern Bayesian approaches, estimate whether particular zeros are sampling or structural zeros and may avoid rarefaction (e.g., [30]). The sampling interval has potentially important effects on the average magnitude of quantitative and qualitative changes in taxon abundances. In the context of unbalanced studies (by design or through missing data), systematic differences in sampling intervals between clinical groups could impact the quantification of microbiome volatility and should be considered as a potential confounder. Despite resulting in a reduction in the total sample size, subsampling pairs of time points to a consistent time interval may serve as a valuable sensitivity analysis.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/article/10.3390/genes14010218/s1, Supplemental Methods: Approach to Subsampling; Figure S1: Selected data and study characteristics for the four studies used in this investigation. Call-out segments correspond to one week of each study, within which sampling frequency is indicated by red circles; Figure S2: Average change in abundance for each taxon at each sampling interval vs. the taxon's relative abundance (small points), with overlaid LOESS smoothers. Changes in taxon abundance are generally centered around 0 regardless of sampling interval and the taxon's average relative abundance. The LOESS curves for the Ravel study show more variability around zero due to the smaller number of nonzero pairs; Figure S3: Intraindividual dissimilarity quantified using four metrics (unweighted UniFrac, generalized UniFrac, weighted UniFrac, and Bray-Curtis dissimilarity) for each time lag in the Moving Pictures study; Figure S4: Distribution of intraindividual Bray-Curtis dissimilarity across rarefaction approaches for each of the four studies; Table S1: Standard deviation (n pairs) of additive changes in taxon relative abundance by study, time lag, and average relative abundance category for each taxon; Table S2: Standard deviation (n pairs) of multiplicative changes between consecutive nonzero relative abundances by study, time lag, and average relative abundance category for each taxon; Table S3: Standard deviation (n pairs) of differences in CLR-transformed abundances between consecutive time points by study, time lag, and average relative abundance category for each taxon; Table S4: Standard deviation (n pairs) of additive changes in relative abundance by study, rarefaction procedure, and average relative abundance category for each taxon; Table S5: Standard deviation (n pairs) of multiplicative changes between consecutive nonzero relative abundances by study, rarefaction procedure, and average relative abundance category for each taxon; Table S6: Standard deviation (n pairs) of differences in CLR-transformed abundances between consecutive time points by study, rarefaction procedure, and average relative abundance category for each taxon.

Author Contributions: Conceptualization, A.M.P.; methodology, software, investigation, and data curation, D.J.P. and A.M.P.; writing—original draft preparation, A.M.P.; writing—review and editing, D.J.P. and A.M.P. All authors have read and agreed to the published version of the manuscript.

Funding: D.J.P.'s work was supported in part by the Roche Student Research Fund at Williams College.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this manuscript are publicly available. Data access, processing, and the analysis code is posted on GitHub at <https://github.com/aplantin/describing-volatility> (manuscript version of code is 9 January 2023). The data processing code, prepared data, subsampling functions, and volatility estimation functions are also available as part of the R package MBVolDescrip in the same GitHub repository.

Conflicts of Interest: AMP declares no conflicts of interest. DJP is employed at Roivant Sciences.

References

1. Cox, S.R.; Lindsay, J.O.; Fromentin, S.; Stagg, A.J.; McCarthy, N.E.; Galleron, N.; Ibraim, S.B.; Roume, H.; Levenez, F.; Pons, N.; et al. Effects of low FODMAP diet on symptoms, fecal microbiome, and markers of inflammation in patients with quiescent inflammatory bowel disease in a randomized trial. *Gastroenterology* **2020**, *158*, 176–188. [[CrossRef](#)] [[PubMed](#)]
2. Golob, J.L.; Pergam, S.A.; Srinivasan, S.; Fiedler, T.L.; Liu, C.; Garcia, K.; Mielcarek, M.; Ko, D.; Aker, S.; Marquis, S.; et al. Stool microbiota at neutrophil recovery is predictive for severe acute graft vs host disease after hematopoietic cell transplantation. *Clin. Infect. Dis.* **2017**, *65*, 1984–1991. [[CrossRef](#)] [[PubMed](#)]
3. Fredricks, D.N.; Plantinga, A.; Srinivasan, S.; Oot, A.; Wiser, A.; Fiedler, T.L.; Proll, S.; Wu, M.C.; Marrazzo, J.M. Vaginal and extra-vaginal bacterial colonization and risk for incident bacterial vaginosis in a population of women who have sex with men. *J. Infect. Dis.* **2022**, *225*, 1261–1265. [[CrossRef](#)] [[PubMed](#)]
4. Vogt, N.M.; Kerby, R.L.; Dill-McFarland, K.A.; Harding, S.J.; Merluzzi, A.P.; Johnson, S.C.; Carlsson, C.M.; Asthana, S.; Zetterberg, H.; Blennow, K.; et al. Gut microbiome alterations in Alzheimer’s disease. *Sci. Rep.* **2017**, *7*, 13537. [[CrossRef](#)]
5. Caporaso, J.G.; Lauber, C.L.; Costello, E.K.; Berg-Lyons, D.; Gonzalez, A.; Stombaugh, J.; Knights, D.; Gajer, P.; Ravel, J.; Fierer, N.; et al. Moving pictures of the human microbiome. *Genome Biol.* **2011**, *12*, R50. [[CrossRef](#)]
6. Chen, L.; Wang, D.; Garmeva, S.; Kurilshikov, A.; Vila, A.V.; Gacesa, R.; Sinha, T.; Segal, E.; Weersma, R.K.; Wijmenga, C.; et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell* **2021**, *184*, 2302–2315. [[CrossRef](#)]
7. Fu, B.C.; Randolph, T.W.; Lim, U.; Monroe, K.R.; Cheng, I.; Wilkens, L.R.; Le Marchand, L.; Lampe, J.W.; Hullar, M.A. Temporal Variability and Stability of the Fecal Microbiome: The Multiethnic Cohort Study Temporal Variability of the Fecal Microbiome. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 154–162. [[CrossRef](#)]
8. Stewart, C.J.; Ajami, N.J.; O’Brien, J.L.; Hutchinson, D.S.; Smith, D.P.; Wong, M.C.; Ross, M.C.; Lloyd, R.E.; Doddapaneni, H.; Metcalf, G.A.; et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **2018**, *562*, 583–588. [[CrossRef](#)]
9. Aho, V.T.; Pereira, P.A.; Voutilainen, S.; Paulin, L.; Pekkonen, E.; Auvinen, P.; Scheperjans, F. Gut microbiota in Parkinson’s disease: Temporal stability and relations to disease progression. *EBioMedicine* **2019**, *44*, 691–707. [[CrossRef](#)]
10. Galloway-Peña, J.R.; Smith, D.P.; Sahasrabhojane, P.; Wadsworth, W.D.; Fellman, B.M.; Ajami, N.J.; Shpall, E.J.; Daver, N.; Guindani, M.; Petrosino, J.F.; et al. Characterization of oral and gut microbiome temporal variability in hospitalized cancer patients. *Genome Med.* **2017**, *9*, 21. [[CrossRef](#)]
11. Schirmer, M.; Denson, L.; Vlamakis, H.; Franzosa, E.A.; Thomas, S.; Gotman, N.M.; Rufo, P.; Baker, S.S.; Sauer, C.; Markowitz, J.; et al. Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. *Cell Host Microbe* **2018**, *24*, 600–610. [[CrossRef](#)] [[PubMed](#)]
12. Gopalakrishnan, V.; Spencer, C.N.; Nezi, L.; Reuben, A.; Andrews, M.; Karpinets, T.; Prieto, P.; Vicente, D.; Hoffman, K.; Wei, S.C.; et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **2018**, *359*, 97–103. [[CrossRef](#)] [[PubMed](#)]
13. Matson, V.; Fessler, J.; Bao, R.; Chongsuwat, T.; Zha, Y.; Alegre, M.L.; Luke, J.J.; Gajewski, T.F. The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **2018**, *359*, 104–108. [[CrossRef](#)] [[PubMed](#)]
14. Bastiaanssen, T.F.; Gururajan, A.; van de Wouw, M.; Moloney, G.M.; Ritz, N.L.; Long-Smith, C.M.; Wiley, N.C.; Murphy, A.B.; Lyte, J.M.; Fouhy, F.; et al. Volatility as a Concept to Understand the Impact of Stress on the Microbiome. *Psychoneuroendocrinology* **2021**, *124*, 105047. [[CrossRef](#)]
15. Halfvarson, J.; Brislawn, C.J.; Lamendella, R.; Vázquez-Baeza, Y.; Walters, W.A.; Bramer, L.M.; D’amato, M.; Bonfiglio, F.; McDonald, D.; Gonzalez, A.; et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* **2017**, *2*, 17004. [[CrossRef](#)]
16. Plantinga, A.M.; Wu, M.C. Beta Diversity and Distance-Based Analysis of Microbiome Data. In *Statistical Analysis of Microbiome Data*; Springer: Cham, Switzerland, 2021; pp. 101–127.
17. Olsson, L.M.; Boulund, F.; Nilsson, S.; Khan, M.T.; Gummesson, A.; Fagerberg, L.; Engstrand, L.; Perkins, R.; Uhlén, M.; Bergström, G.; et al. Dynamics of the normal gut microbiota: A longitudinal one-year population study in Sweden. *Cell Host Microbe* **2022**, *30*, 726–739. [[CrossRef](#)]
18. Vandeputte, D.; De Commer, L.; Tito, R.Y.; Kathagen, G.; Sabino, J.; Vermeire, S.; Faust, K.; Raes, J. Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nat. Commun.* **2021**, *12*, 6740. [[CrossRef](#)]
19. Mehta, R.S.; Abu-Ali, G.S.; Drew, D.A.; Lloyd-Price, J.; Subramanian, A.; Lochhead, P.; Joshi, A.D.; Ivey, K.L.; Khalili, H.; Brown, G.T.; et al. Stability of the human faecal microbiome in a cohort of adult men. *Nat. Microbiol.* **2018**, *3*, 347–355. [[CrossRef](#)]
20. Plantinga, A.M.; Chen, J.; Jenq, R.R.; Wu, M.C. pldist: Ecological dissimilarities for paired and longitudinal microbiome association analysis. *Bioinformatics* **2019**, *35*, 3567–3575. [[CrossRef](#)]
21. Silverman, J.D.; Durand, H.K.; Bloom, R.J.; Mukherjee, S.; David, L.A. Dynamic linear models guide design and analysis of microbiota studies within artificial human guts. *Microbiome* **2018**, *6*, 202. [[CrossRef](#)]
22. Shenhav, L.; Furman, O.; Briscoe, L.; Thompson, M.; Silverman, J.D.; Mizrahi, I.; Halperin, E. Modeling the temporal dynamics of the gut microbial community in adults and infants. *PLoS Comput. Biol.* **2019**, *15*, e1006960. [[CrossRef](#)] [[PubMed](#)]
23. Sinha, R.; Goedert, J.J.; Vogtmann, E.; Hua, X.; Porras, C.; Hayes, R.; Safaeian, M.; Yu, G.; Sampson, J.; Ahn, J.; et al. Quantification of human microbiome stability over 6 months: Implications for epidemiologic studies. *Am. J. Epidemiol.* **2018**, *187*, 1282–1290. [[CrossRef](#)] [[PubMed](#)]

24. Flores, G.E.; Caporaso, J.G.; Henley, J.B.; Rideout, J.R.; Domogala, D.; Chase, J.; Leff, J.W.; Vázquez-Baeza, Y.; Gonzalez, A.; Knight, R.; et al. Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **2014**, *15*, 531. [[CrossRef](#)] [[PubMed](#)]
25. Gajer, P.; Brotman, R.M.; Bai, G.; Sakamoto, J.; Schütte, U.M.; Zhong, X.; Koenig, S.S.; Fu, L.; Ma, Z.; Zhou, X.; et al. Temporal dynamics of the human vaginal microbiota. *Sci. Transl. Med.* **2012**, *4*, 132ra52. [[CrossRef](#)] [[PubMed](#)]
26. Ravel, J.; Brotman, R.M.; Gajer, P.; Ma, B.; Nandy, M.; Fadrosh, D.W.; Sakamoto, J.; Koenig, S.S.; Fu, L.; Zhou, X.; et al. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome* **2013**, *1*, 29. [[CrossRef](#)]
27. Martín-Fernández, J.A.; Barceló-Vidal, C.; Pawlowsky-Glahn, V. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math. Geol.* **2003**, *35*, 253–278. [[CrossRef](#)]
28. Martín-Fernández, J.A.; Hron, K.; Templ, M.; Filzmoser, P.; Palarea-Albaladejo, J. Model-based replacement of rounded zeros in compositional data: Classical and robust approaches. *Comput. Stat. Data Anal.* **2012**, *56*, 2688–2704. [[CrossRef](#)]
29. Martín-Fernández, J.A.; Hron, K.; Templ, M.; Filzmoser, P.; Palarea-Albaladejo, J. Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat. Model.* **2015**, *15*, 134–158. [[CrossRef](#)]
30. Liu, T.; Zhao, H.; Wang, T. An empirical Bayes approach to normalization and differential abundance testing for microbiome data. *BMC Bioinform.* **2020**, *21*, 225. [[CrossRef](#)]
31. Mandal, S.; Van Treuren, W.; White, R.A.; Eggesbø, M.; Knight, R.; Peddada, S.D. Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **2015**, *26*, 27663. [[CrossRef](#)]
32. Chen, J.; Bittinger, K.; Charlson, E.S.; Hoffmann, C.; Lewis, J.; Wu, G.D.; Collman, R.G.; Bushman, F.D.; Li, H. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **2012**, *28*, 2106–2113. [[CrossRef](#)] [[PubMed](#)]
33. McMurdie, P.J.; Holmes, S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **2014**, *10*, e1003531. [[CrossRef](#)] [[PubMed](#)]
34. Hong, J.; Karaoz, U.; de Valpine, P.; Fithian, W. To rarefy or not to rarefy: Robustness and efficiency trade-offs of rarefying microbiome data. *Bioinformatics* **2022**, *38*, 2389–2396. [[CrossRef](#)] [[PubMed](#)]
35. Cameron, E.S.; Schmidt, P.J.; Tremblay, B.J.M.; Emelko, M.B.; Müller, K.M. To rarefy or not to rarefy: Enhancing diversity analysis of microbial communities through next-generation sequencing and rarefying repeatedly. *Sci. Rep.* **2021**, *11*, 22302. [[CrossRef](#)] [[PubMed](#)]
36. Willis, A.D. Rarefaction, alpha diversity, and statistics. *Front. Microbiol.* **2019**, *10*, 2407. [[CrossRef](#)] [[PubMed](#)]
37. Ramakodi, M.P. Effect of amplicon sequencing depth in environmental microbiome research. *Curr. Microbiol.* **2021**, *78*, 1026–1033. [[CrossRef](#)] [[PubMed](#)]
38. Tsilimigras, M.C.; Fodor, A.A. Compositional data analysis of the microbiome: Fundamentals, tools, and challenges. *Ann. Epidemiol.* **2016**, *26*, 330–335. [[CrossRef](#)] [[PubMed](#)]

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.