ABSTRACT

Many microbial ecology experiments use sequencing data to measure a community's response to an experimental treatment. In a common experimental design, two units, one control and one experimental, are sampled before and after the treatment is applied to the experimental unit. The four resulting samples contain information about the dynamics of organisms that respond to the treatment, but there are no analytical methods designed to extract exactly this type of information from this configuration of samples. Here we present an analytical method specifically designed to visualize and generate hypotheses about microbial community dynamics in experiments that have paired samples and few or no replicates. The method is based on the Poisson lognormal distribution, long studied in macroecology, which we found accurately models the abundance distribution of taxa counts from 16S rRNA surveys. To demonstrate the method's validity and potential, we analyzed an experiment that measured the effect of crude oil on ocean microbial communities in microcosm. Our method identified known oil degraders as well as two clades, *Maricurvus* and Rhodobacteraceae, that responded to amendment with oil but do not include known oil degraders. Our approach is sensitive to organisms that increased in abundance only in the experimental unit but less sensitive to organisms that increased in both control and experimental units, thus mitigating the role of "bottle effects".

INTRODUCTION

Paired-sample microbial ecology experiments

Many microbial ecology experiments use amplicon-based sequencing (e.g., 16S rRNA gene sequencing) to study the dynamics of a microbial community, whose members are typically grouped into operational taxonomic units (OTUs). A straightforward experimental design uses control units (i.e., subjects, microcosms, or animals not subjected to the treatment) and pretests (i.e., measurements of microbial community composition taken before the treatment is applied to the experimental units). This two-timepoint, paired-sample design is intended to identify changes in community composition that are specific to the experimental treatment, rather than all the changes that occur in the experimental unit, some of which might be irrelevant to the intentionally applied treatment. For example, aquatic microbial communities transplanted into microcosms often suffer dramatic changes in community composition on account of the transplantation. These extraneous community composition changes are called "bottle effects". Over the course of the experiment, bottle effects are entangled with the effects caused by the experimental treatment [1]. A similar complication can occur in experiments that study animal-associated microbiota, where changes in community composition can be caused by, for example, interactions with experimenters or circulation of a microbe in the animal facility.

If an experiment is performed with many replicates, there are statistical methods that can make statements about whether the observed changes between the groups were meaningful with respect to a null hypothesis. In some cases, an experiment is very

preliminary or the samples are so precious that robust replication is not feasible. For example, we intended the test the effect of crude oil on an ocean microbial community and did not have enough of the sample material, which was collected on a special cruise, for many replicates. Without sufficient replications, it is impossible for us to make a statement about the statistical significance of the results of such an experiment, but we wanted to obtain as much information as possible from such difficult-to-acquire samples, with the caveat that any conclusion would need a separate, properly-powered experiment for verification.

Because of the complications in paired-sample experiments in general and our experimental setup in particular, we aimed to create a method to visualize and generate hypotheses about the dynamics of microbial communities in paired-sample experiments with few or no replicates. Most popular tools for analyzing the results of microbial ecology experiments have shortcomings or complications when applied to this particular experimental setup. We review some classes of these existing tools below.

For convenience, we refer to the four samples in a two-timepoint, two-unit experiment as "control-before" (control unit, pretest sample), "control-after" (control unit, post-treatment sample), "experimental-before", and "experimental-after".

Existing analytical techniques

Ordination techniques. Ordination techniques include principal component analysis (PCA), multidimensional scaling (MDS), redundancy analysis (RDA), and canonical

analysis of principal components (CAP) [2, 3]. Although very useful for analyzing and visualizing the relationships between many samples, ordination techniques are not easy to interpret in the context of a paired-sample microbial ecology experiment. In an experiment with no replicates, an ordination plot will only have four points, making it difficult to visually or analytically extract interesting information about the community's dynamics.

Clustering techniques. In the context of microbial ecology, clustering techniques like hierarchical clustering or *k*-means clustering use a dissimilarity metric to infer which samples in a data set group together [2]. Clustering techniques are often combined with ordination techniques and give similar insight. They therefore suffer similar drawbacks when analyzing the results of paired-sample experiments. In the case of a single paired-sample experiment with no replicates, there are only four samples, so a clustering technique can only produce a limited number of analytical outcomes.

Beta diversity and other tests. Statistical tests like ANOSIM [4] and PERMANOVA [5] are designed to evaluate whether some set of samples in a data set are more similar to one another than they are to other samples in the data set. In general, rigorous statistical inference and machine learning methods require many replicates, which might be impractical for exploratory studies. For example, La Rosa *et al*. [6] calculate that at least 25 samples are needed to determine if two groups of human microbiome samples have meaningfully different compositions at 90% power.

Indicator species techniques. Indicator species techniques like IndVal [7] are intended to identify species that especially informative with respect to the ecological community's composition or abiotic context [8]. Indicator species techniques treat each sample as an independent community rather than, as we consider, timepoints from the same community. For example, if an OTU is rare in the control-before and control-after but abundant in the experimental-before and experimental-after, an indicator species approach would identify that OTU as important to distinguishing the experimental and control *units*, which, although true, does not correctly reflect that that OTU is probably not affected by the experimental treatment, since it changed in neither the control or experimental unit.

OTU-by-OTU techniques. There are techniques that can identify OTUs whose dynamics merit further investigation, even if those OTUs are not abundant enough to appear in a community composition chart or if there are not enough replicates for statistical inference. For example, OTUs can be ordered according to their change in relative abundance between the pretest and the post-treatment sample. However, unintuitive signals, called "compositional effects", can arise when standard analytical techniques are used with compositional data sets like OTU count data [9, 10]. In an aquatic microcosm experiment, for example, a single organism might bloom in response to the experimental treatment, causing an organism with constant absolute abundance to decrease in relative abundance. The reverse is also possible: an organism with constant absolute abundance increases in relative abundance when other organisms decrease in absolute abundance.

Opportunity for a distribution-based technique

To incorporate controls and pretests, it is important to be able to meaningfully compare OTUs' abundances across samples. As discussed above, relative abundances are susceptible to compositional effects. In contrast, nonparametric analysis, which uses only the ranks of abundance-ordered OTUs, is robust to compositional effects but loses much of the quantitative information encoded in the relative abundances. For example, an arbitrarily large change in the abundance of one OTU can cause arbitrarily large changes in the relative abundances of all OTUs but will only change the ranks of each OTU by, at most, one. Using ranks presents a tradeoff between robustness (i.e., each rank changing by at most one) and loss of quantitative information (e.g., if the most abundant OTU doubles in abundance, no ranks change). Ranks are also challenging to use with OTU count data because many OTUs have the same number of counts.

If OTU abundances were distributed in a way that could be reliably modeled, a compromise between relative abundances and ranks would be possible. For example, if one organism blooms and all others remain at constant absolute abundance, the overall shape of the distribution of abundances would change very little. As in an analysis that uses relative abundances, the blooming OTU would be registered because that OTU would move up in the distribution. As in an analysis that uses ranks, the unchanging OTUs would remain at the same places in the distribution even though their relative abundances decreased.

Our method

In this paper, we present an analytical method designed to measure the dynamics of OTUs across two timepoints, to correct for bottle effects using control units, and to correct for unit-specific effects using pretests. The method is framed in terms of an abundance distribution from macroecology research, the Poisson lognormal distribution, that we found accurately models the abundance distribution of OTU count data. As a test of the method's validity, we show that, in the context of a bioreactor experiment, this method reports that OTUs in microbial communities derived from the same inoculum and subjected to strong but identical conditions have well-correlated responses.

We used our method to identify OTUs in a complex ocean water community, collected off the Egyptian coast, that respond to amendment with crude oil. Most of the sequences we identified classified as *Maricurvus*, *Pseudomonas*, *Alcanivorax*, *Methylophaga*, and Rhodobacteraceae. These clades include known oil degraders as well as organisms that other microbial ecology experiments have suggested may degrade oil. These results demonstrate that our method will be useful for visualizing the effect of the treatment of interest on all OTUs and for quantifying dynamic changes in abundance in a paired-sample microbial ecology with few or no replicates, although properly-powered follow-up experiments would be needed to verify any of these dynamics.

MATERIALS & METHODS

Poisson lognormal distribution

Theory. Noting that many microbial communities are structured by complex ecological processes, we searched for an ecologically-motivated probability distribution that accurately models the abundance distribution of OTUs in natural microbial communities. The truncated Poisson lognormal (TPL) distribution is an attractive candidate. When a value is drawn from the TPL distribution, a true abundance λ is first drawn from a lognormal distribution (with scale parameter μ and shape parameter σ). Then a random integer is drawn from a Poisson distribution with mean λ. If the integer is zero, a new λ is drawn and is used to draw a new integer.

Among the many models of species-abundance relationships (e.g., [11, 12] and references therein), there is evidence and theory suggesting that fractions of the total niche space allotted to each organisms are approximately lognormal-distributed [13], and the Poisson distribution is a straightforward model for converting a continuous value λ into a random number of discrete counts. The Poisson lognormal has been used to model abundance distributions for plants and animals [14-16] and has been used in at least one study [17] that simulated microbial abundances. In most of these applications, the distribution is truncated at zero counts, since, in most cases, it is impossible to distinguish if a species is absent or if it present but very rare; in both cases, that species would present zero counts.

In a microbial ecology context, the TPL framework asserts that the abundances of

microbial species in an environment are lognormal-distributed, that is, that the logarithms of those abundances have a Gaussian distribution. The framework also asserts that sequencing produces an integer number of reads for each species. The number of sequencing counts for a species with true abundance λ is drawn from a Poisson distribution with mean λ. The parameter μ is related to the mean of the abundances of microbial species in that environment (conditioned on the depth of sequencing). The parameter σ describes the variability of those abundances.

Metrics. Once fit to the OTU abundances in a sample, the TPL distribution provides transformations from raw OTU counts to two different values. As mentioned above, in the TPL framework, OTUs' true abundances λ are assumed to be lognormal-distributed with scale μ and shape σ. If the Poisson lognormal distribution is an accurate model for OTU abundance distributions, samples might have differing parameters μ and σ, but the underlying distribution of $(\log \lambda - \mu)/\sigma$ should be similar across samples. We expect this because if, roughly speaking, λ is lognormal distributed, then $\log \lambda$ is normally distributed, and $(\log \lambda - \mu)/\sigma$ accounts for the differences in means and shapes of the normal distribution. An OTU's number of reads r is the maximum likelihood estimator of its true abundance λ, so we define the *normalized reads* $z \equiv (\log r - \mu)/\sigma$, which estimates $(\log \lambda - \mu)/\sigma$.

Given this metric z that can be compared across samples, we looked for a sensible way to combine these values as a measure of dynamics. To quantify an OTU's dynamics between the two timepoints, we define Δz, the change in rescaled reads. This metric is

similar to the log fold change in relative abundance, an application of the more general log-ratio transformation commonly used with compositional data sets. To show this connection, consider two samples $i \in \{0 = \text{before}, 1 = \text{after}\}$ from one microcosm, either control or experimental, with the TPL fit parameters μ_i and σ_i. One OTU of interest has counts r_i in the two samples. In this case,

$$\Delta z \equiv \frac{\log r_1 - \mu_1}{\sigma_1} - \frac{\log r_0 - \mu_0}{\sigma_0} = \frac{\log r_1}{\sigma_1} - \frac{\log r_0}{\sigma_0} + \text{constant}.$$

For comparison, the log fold change in relative abundance is

$$\log\left(\frac{r_1/N_1}{r_0/N_0}\right) = \log r_1 - \log r_0 + \text{constant},$$

where N_i is the total number of reads in sample i. If $\sigma_0 \approx \sigma_1$, then

$$\Delta z \approx \frac{\log \text{ fold change}}{\sigma_0} + \text{constant},$$

that is, Δz and the log fold change are approximately linearly related.

Aside from rescaled reads, the TPL distribution can also be used to transform raw OTU counts to the value of the cumulative distribution function F, which has the common range [0, 1] across all samples. Having fit the parameters μ and σ, an OTU with r reads has $F(r) \equiv \sum_{r' \leq r} f_{\text{TPL}}(r')$, where f_{TPL} is the probability distribution function for the TPL distribution. Conveniently, ΔF is defined for all r_0 and r_1, while Δz and the log fold change are poorly defined when $r_0 = 0$ or $r_1 = 0$.

Implementation. We implemented the TPL fitting, computation of z and F, and basic visualizations in an R package *texmexseq*, which we have posted at CRAN (cran.r-

project.org). Our package is based on the poilog package [18], to which we add convenience functions for interacting with OTU tables, fitting the distribution to multiple samples, and visualizing Δz and ΔF.

Bioreactor experiment

Experimental design. We performed a bioreactor experiment to verify that our analytical method would identify similar dynamics in identically-treated microcosms. Briefly, three identical anaerobic serum bottles were loaded with sterile anaerobic media, glucose, and anaerobic sludge from Al Mafraq wastewater treatment plant (Abu Dhabi Sewage Services Company, Al Dhafrah, Abu Dhabi, UAE). Glucose was the only carbon source. Each bioreactor was incubated at 35 °C for 48 hours, at which point the bioreactor's contents was spun down and the cell pellet resuspended in fresh media. This process was repeated 7 times. The timepoints analyzed in this study are the initial inoculum's community and the bioreactor's final community (i.e., timepoint 7).

Experimental protocol. Inocula were stored at 4 °C before starting experiments. Fermentations were carried out in 150 mL serum bottles with a working volume of 60 mL. Anaerobic serum bottles were loaded with media, described below, and sludge. The initial biomass concentration for all the three inocula was 10 g/L dry weight matter. Sterile media consisted of 5 g/L glucose (autoclaved separately from mineral media) and phosphate buffer (0.2 g/L $Na_2HPO_4 \cdot 2H_2O$ and 2.5 g/L KH_2PO_4) diluted on basal anaerobic media [19]. Media pH was adjusted to 5.5 with 1 M HCl.

After inoculation, each bottle was crimped using sterile rubber stoppers and flushed with pure N_2 for 2 minutes using sterile 0.45 μm pore gas filters. Bottles were incubated immediately after flushing. To re-suspend the inoculum, the broth was centrifuged in sterilized containers at 5,000 g for 5 minutes. The resulting pellets were re-suspended in 60 mL of fresh media. Bottles were again crimped and flushed. Inoculation and media replacement were all performed in a UV-sterilized laminar flow chamber.

DNA from the bioreactors was extracted with MO BIO Ultra Clean Soil DNA isolation kit according to the manufacturer's protocol. Paired-end Illumina sequencing libraries were constructed using a two-step 16S rRNA PCR amplicon approach described in Preheim *et al.* [20]. Libraries were multiplexed and sequenced on an Illumina MiSeq with paired-end 150 bp reads.

16S data processing. Only forward reads were used in the analysis. Primers were trimmed from the reads by searching for the best-matching position in the read's first 20 bases, allowing a maximum of 2 mismatches between the primer sequence and the read sequence. Reads that did not match the primer were discarded. Reads were demultiplexed by assigning reads to the best-matching barcode sequence, allowing no more than 1 mismatched base. Reads with no acceptable barcode match were discarded. Reads were trimmed to 120 bases. Shorter reads were discarded. Reads with an expected number of errors, as calculated by Edgar & Flyvbjerg [21], greater than 1.0 were discarded. The sequence data for these experiments are in Datasets S1 and S2.

Aquatic microcosm experiment

Inoculum collection. Water samples were collected on 12/13 October 2012 aboard MV *Fugro Navigator* in the West Nile Delta region of the Nile Deep Sea Fan from a station (29.571° E, 31.813° N) with a sea floor depth of 1230 m. This work was conducted in BP's West Nile Delta Concession as part of a larger survey of Eastern Mediterranean ocean microbiology described elsewhere [22]. No specific permits were required for collection of these samples. These field studies did not include the collection of any endangered or protected species. Temperature, salinity, depth, and dissolved oxygen were measured through the water column with a Valeport Midas+ CTD. Samples were collected with Niskin bottles from four depths selected in consideration of differences in temperature, salinity, and depth: within the thermocline (50 m), within an area of increased salinity in the water column (250 m), two-thirds of sea floor depth (824 m), and 20 m above the sea floor (1210 m). Water was removed from the Niskin bottles and stored in pre-cleaned amber glass bottles at 4 °C until the microcosms were set up. In this paper, we analyze the result of the experiment performed using water from 824 meters depth.

Microcosm design and sampling. We performed a microcosm experiment to evaluate the effect of crude oil on the microbial community in those water samples. 2 L of water from each depth was used for microcosms, 1 L each for a control microcosm and experimental microcosm. Microcosms were incubated at room temperature in amber glass bottles wrapped in tin foil. The experimental microcosms were treated with 100 ppm v/v crude oil. The oil, Norne Blend, was selected because we expected it would be similar in

composition to oil in natural reservoirs near the sampling site. When we sampled, there were no wells near the sampling site that were producing oil.

At 0 and 72 hours after the microcosms were prepared, 100 mL subsamples were extracted and immediately filtered through a 0.2 μm filter. DNA was extracted from the filters according to the standard protocol for the MoBio PowerWater DNA Isolation Kit. We amplified a subunit of the V4 region of the 16S rRNA gene following the procedure described in Preheim *et al.* [20]. Extracted DNA was amplified using custom barcoded primers and sequenced with paired-end 250 bp reads on an Illumina MiSeq.

16S data processing. Primers were trimmed from the forward and reverse reads by searching for the best-matching position in the read's first 20 bases, allowing a maximum of 2 mismatches between the primer sequence and the read sequence. Read pairs without matching forward and reverse primers were discarded. Reads were demulitplexed by assigning reads to the best-matching barcode sequence, allowing no more than 1 mismatched base. Reads with no acceptable barcode match were discarded. Reads were merged by (i) evaluating alignments that would produce merged reads of 263 ± 5 bases, (ii) selecting the alignment with the greatest number of matching bases, (iii) assigning consensus bases and quality scores according to Edgar & Flyvbjerg [21]. Merged reads with more than 2.0 expected errors were discarded. Taxonomic information was collected using the RDP classifier [23] and, in select cases, NCBI BLAST [24]. We searched for chimeras in the data by performing UCHIME with the Broad gold database [25], UCHIME with the RDP training database (version 9), and *de novo* UPARSE at 99%

identity [26]. The sequence data for these experiments are in Datasets S3 and S4 and at MG-RAST [27] under accession 4685010.3.

OTU calling. For most analysis of the ocean and sludge experiments, we use unique sequences as OTUs (i.e., these are 100% identity OTUs). For visual clarity, we used 99% *de novo* clustering with UPARSE for all samples in Figure 2. In Figure S1, we called OTUs using multiple methods. To call OTUs with RDP, we truncated every sequence's taxonomy at the first level that has less than 80% bootstrap confidence, and two sequences that have the same truncated taxonomy are placed in the same OTU. To call reference-based OTUs, we used the Greengenes reference database [28] (August 2013 97% OTUs) and global usearch [29] with minimum 97% identity.

Phylogenetic tree. The sequences shown on the tree are the 303 sequences most abundant in the four microcosm experiments. Sequences were aligned to the Greengenes core set [28] using Pynast [30]. One sequence (#229) that did not align to the core set was excluded. The tree was constructed with FastTree [31] and drawn with APE [32].

HMP samples

The Human Microbiome Project [33] samples were downloaded from the 16S rRNA trimmed data set (HM16STR). Reads from the stool sample (#700014956) and vaginal sample (#700016101) were trimmed to 200 bp. Only V3-V5 region reads were used.

RESULTS

Disparate dynamics in control and experimental microcosms

Figure 1 shows data from our paired-sample aquatic microcosm experiment. These data highlight some of the issues mentioned in the Introduction. One of the OTUs (OTU 3) increases to a high abundance in the experiment-after sample, potentially introducing a large compositional effect. Furthermore, that OTU's apparently dramatic dynamics in the experimental microcosm should be treated with skepticism because OTU 3 also increases in the control unit. In contrast, another OTU (OTU 63) was not detected in the two control samples or in the experiment-before sample, but it has a high abundance in the experiment-after sample, suggesting that little or none of its increase in the experimental microcosm should be attributed to bottle effects. How should we compare an OTU's abundances in the four samples? How should we correct results of the experimental unit with information from the control unit? We were motivated to develop a method based on the Poisson lognormal distribution to answer these questions.

Poisson lognormal distribution accurately models 16S abundances

We found that the truncated Poisson lognormal (TPL) distribution is an excellent fit for the abundance distributions of OTUs from multiple environments (Figure 2). To quantify the quality of the fit, we conducted an empirical test to see if the differences between the theoretical and observed abundances can be attributed to chance. For each sample shown in Figure 2, we fit the Poisson lognormal distribution to the sample's data, simulated 10 000 datasets (each with a number of OTUs equal to the number in the observed data) using the fit parameters, and compared the chi-square goodness-of-fit statistic in these

simulations to the goodness-of-fit in the observed data. In all cases, the differences

between the theoretical and observed distributions are attributable to chance ($p = 0.76$,

0.91, 0.66 for the first three panels in Figure 2), although the attribution to chance was

marginal in the sample obtained from a microcosm after treatment with oil ($p = 0.054$;

"oiled ocean" in Figure 2). The TPL distribution's fit is similar when the OTUs were

called with some other common OTU-calling methods (Figure S1).

Replicate units yield well-correlated results

To check that the TPL distribution can be used to quantify dynamics, we compared

replicates from the bioreactor experiment, which subjects a microbial community to

strong selective pressures. We expected that strong selective pressures would cause

dramatic changes in microbial community composition but that these effects would be

similar across replicates. Analytically, this means that we expect that Δz, a measure of an

OTU's dynamics, should be similar across replicates. Visually, this means that, if the Δz

values for all OTUs in two replicates are plotted against one another, they should fall

along the $y = x$ diagonal. In fact, the replicated bioreactors show these sorts of well-

correlated dynamics (Figure 3). Each plot provides an immediate summary of the

relationship between the dynamics of all OTUs in the four samples in the experiment.

Identification of known and putative oil-degrading organisms

Having shown that metrics derived from the TPL distribution can be used to quantify

dynamics, we analyzed the results of an aquatic microcosm experiment with one control

microcosm, one experimental microcosm, and two timepoints (pretest and post-treatment

samples). In this experiment, ocean water was treated with crude oil to gain insight into the effects of a potential oil spill in this region. Previous work has shown that, in many aquatic environments, a few species (especially in the genera *Alcanivorax* and *Cycloclasticus*) multiply to make up the majority of microbes after crude oil is added [34, 35]. We aimed to identify OTUs in this ecosystem that respond to amendment with crude oil.

To identify OTUs with suggestive responses to oil, we selected criteria for Δz and ΔF that would be consistent with a response to oil and not growth due to bottle effects. Specifically, we considered OTUs whose Δz in the experimental unit was greater than or equal to its Δz in the control unit minus 0.5. Roughly speaking, this selected OTUs whose abundances increased more in the experimental unit relative to the control unit. We selected this cutoff criterion because it was compatible with our expectations about specific responses to oil and also included some OTUs with finite Δz values (Figure S2). Analogous to the fold-change cutoffs used to identify interesting spots on microarrays, this sort of cutoff criterion does not necessarily OTUs whose dynamics would be considered statistically significant in a well-powered experiment.

We also considered OTUs whose ΔF in the treatment unit was greater than 0.5 but whose ΔF in the control unit was less than 0.5. Roughly speaking, this selected for OTUs whose position in the TPL distribution moved *up* past about 50% of other OTUs in the experimental unit but whose position in the distribution moved *down* past about 50% of OTUs in the control unit. A plot of the ΔF values which highlights the OTUs that meet

the criteria are shown in Figure S3. The OTUs that satisfy either the Δz or ΔF criteria are shown in the context of a bacterial phylogeny in Figure 4. They group into five monophyletic clades. Detailed information about the OTU's taxonomic classification and dynamics are reported in Table 1.

The OTUs that appear to respond to oil are all members of γ-Proteobacteria (clades A through D) or α-Proteobacteria (clade E). Among the γ-Proteobacteria, many of these OTUs correspond to phylogenetic groups that contain known oil degraders: the two OTUs in clade C are both classified as *Alcanivorax*, the seven OTUs in clade B are classified as *Pseudomonas* or Pseudomonadaceae [36], and the one OTU in clade D is classified as *Methylophaga* [37, 38].

All but one of the eight OTUs in clade A are classified as *Maricurvus*. These seven OTUs align to NCBI entries for *Maricurvus nonylphenolicus* and *Aestuariicella hydrocarbonica*. The first species, *M. nonylphenolicus* is the *Maricurvus* type strain and degrades nonylphenol [39], while the second, *A. hydrocarbonica* has a 16S sequence highly similar to *Maricurvus* and degrades multiple aliphatic hydrocarbons [40].

The OTUs in clade E, which are members of α-Proteobacteria, are classified by RDP as Rhodobacteracea, and they align equally well to 16S sequences from *Phaeobacter*, *Roseobacter*, *Pelagimonas*, and *Sulfitobacter* spp. Although genus *Phaeobacter* has no known oil-degrading species, it may have increased in abundance in other experiments that amended ocean water with crude oil [41], and *Sulfitobacter* spp. were abundant in

oiled beach sands [35] and in a large microcosm simulating an oil spill in ocean water

[42].

DISCUSSION

The truncated Poisson lognormal distribution and microbial ecology

As noted above, the TPL distribution has been used to model the abundance distribution of plants and animals, but to our knowledge this is the first report in which the TPL distribution is used to model microbial abundances collected in 16S surveys.

When we laid out the logic of the TPL distribution, we described the Poisson distribution as the link from the continuous-valued true abundance λ to the discrete number of sequencing reads. However, the number of 16S genes is not identical between organisms, and so our approach uses a single layer (a stochastic change from λ to the number of reads) to model a process that actually has two layers (a deterministic change from the organism's true abundance λ to the true abundance of its 16S gene, and from there to reads). The quality of the TPL distribution's fit to 16S abundance data suggests that variations in 16S copy number need not be separately included to explain the observed abundance distributions. This idea contrasts against the approach of Kembel *et al.* [17], who showed that the abundance distribution of organismal counts N, computed from an estimate of a taxon's 16S copy number C and its 16S sequence data counts $N \times C$, fits the lognormal better than does the abundance distribution of 16S counts $N \times C$. Our results suggest that the lognormal is simply a poor fit for discrete 16S count data. In the ecosystems we studied, compounding the lognormal with a random Poisson distribution is sufficient to make an excellent fit to 16S OTU abundance data.

Interpreting microbial dynamics

The information in Table 1 helps demonstrate that three quantifications of OTU dynamics—change in relative abundance, change in rescaled reads z, and change in cumulative distribution function value F—provide complementary information about the OTUs' dynamics in our experiment. In multiple cases, OTUs satisfied both the Δz and ΔF criteria. However, the ΔF criteria tend to identify organisms that are less abundant and have non-finite Δz values. Some OTUs underwent a small change in relative abundance but a large change in ΔF, indicating the abundance distribution for 16S sequences is so skewed that that small change in relative abundance is sufficient for it to advance dramatically relative to other OTUs. For example, OTU 63 in clade A experiences a small increase in relative abundance (+0.8%) in the experimental microcosm, but its ΔF (+0.996 of a possible 1.0) indicates that its small change in relative abundance made it more abundant than most of the other OTUs in the sample. In this case, a small change in relative abundance for a lowly-abundant OTU can equate to a large ΔF. Conversely, a large change in relative abundance for a highly-abundant OTU can equate to a small ΔF.

We speculate that the Poisson lognormal's fit to microbial community structure may have further relevance to making inferences about microbial community dynamics. The problem of limited replicates is not new, and difficulties in replication were probably more pronounced for microarray studies. One approach to limited replication in microarray studies was to infer the variance in each gene's expression level, potentially improving the power of gene-by-gene t-tests without requiring more replicates, using Bayesian hierarchical models [43]. Just as genes with higher expression level tend to

have higher variance, it may be that there are trends in the variance or dynamics of OTUs' relative abundances. It may be that the Poisson lognormal distribution could provide a lens for discovering those trends, which in turn might provide better inference for uncovering OTUs' dynamics if they are integrated in a productive way. It is important to again note that no method, no matter how sophisticated, can obviate the need for sufficient replication.

As mentioned in the Introduction, sequencing count data sets are compositional and are therefore subject to compositional effects. Without a quantification of overall or absolute bacterial abundance, metagenomic 16S datasets can provide evidence —but never proof— that a measured bacterial species changed in absolute abundance. However, if the Poisson lognormal distribution accurately models the distribution of OTUs' abundances, and if most OTUs do not change in absolute abundance, then the OTUs' positions with respect to the distribution may reflect their absolute abundances better than their relative abundances do. A well-powered experiment comparing absolute abundances and OTU's positions in the Poisson lognormal distribution could evaluate this possible relationship.

Possible identification of oil degraders

A simple microcosm experiment, coupled with our analytical technique, identified many OTUs that may have increased meaningfully in abundance in response to the amendment with crude oil. OTUs classified as Rhodobacteraceae had Δz values indicating a small bloom, suggesting that these organisms might be involved in oil degradation without being capable of degrading oil on their own. On the other hand, the large blooms of

OTUs classified as *Maricurvus*, along with the recent discovery of the closely-related oil degrader *Aestuariicella hydrocarbonica*, suggest that these organisms might be relevant to oil degradation in the Eastern Mediterranean.

Uses and limitations for this method

In our experiment, we had enough sample material for a few microcosms, and we could non-destructively subsample each microcosm through time. Ideally, we would have had access to much more sample material so we could run many replicate experiments and use standard statistical approaches to validate any observed bacterial community dynamics. In our constrained setup, we aimed to mitigate extraneous, confounding effects like time, bottle effects, or contact with a common set of microbes by incorporating information about the community composition dynamics from multiple timepoints. Experiments in which subsampling is less onerous than replications —as in, for example, experiments studying the microbiome of animals in a facility or experiments studying the effect of a treatment *in situ*— might benefit from quantitative correction of experimental results using the data obtained from pretests and control units.

We expect that this method might also be applicable to situations beyond the very constrained one that originally motivated us. If an experimental setup has many replicates, it may be that performing rigorous statistical analyses on the Poisson lognormal metrics, rather than just the relative abundances, yields more useful results. We also expect that the Poisson lognormal distribution might fit data from amplicon-based sequencing of other taxonomic marker genes, like eukaryotes' 18S rRNA gene.

ACKNOWLEDGEMENTS

Data Availability: All relevant data are within the paper and its Supporting Information files. The oil microcosm sequences are also available at MG-RAST under accession 4685010.3.

SUPPORTING DATASET LEGENDS

Dataset S1. Unique sequences from sludge bioreactor experiments. Trimmed, quality filtered, dereplicated sequences.

Dataset S2. Count data from sludge bioreactor experiments. Table showing number of times each sequence appeared in each sample (three replicates, two timepoints each).

Dataset S3. Unique sequences from oil microcosm experiments. Trimmed, merged, quality filtered, dereplicated sequences.

Dataset S4. Count data from oil microcosm experiments. Table showing number of times each sequence appeared in each sample (water from four depths, control and experimental "oil" microcosms, two timepoints each).

REFERENCES

1.   Hammes F, Vital M, Egli T. Critical Evaluation of the Volumetric "Bottle Effect" on Microbial Batch Growth. Appl Environ Microbiol. 2010;76(4):1278-81. doi: 10.1128/aem.01914-09.

2.   Ramette A. Multivariate analyses in microbial ecology. FEMS Microbiol Ecol. 2007;62(2):142-60. doi: 10.1111/j.1574-6941.2007.00375.x.

3.   Anderson MJ, Willis TJ. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology. 2003;84(2):511-25. doi: 10.1890/0012-9658(2003)084[0511:CAOPCA]2.0.CO;2.

4.   Clarke KR. Non-parametric multivariate analyses of changes in community structure. Austral Ecol. 1993;18(1):117-43. doi: 10.1111/j.1442-9993.1993.tb00438.x.

5.   Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32-46. doi: 10.1111/j.1442-9993.2001.01070.pp.x.

6.   La Rosa PS, Brooks JP, Deych E, Boone EL, Edwards DJ, Wang Q, et al. Hypothesis testing and power calculations for taxonomic-based human microbiome data. PLoS ONE. 2012;7(12):e52078. doi: 10.1371/journal.pone.0052078.

7.   Dufrene M, Legendre P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol Monogr. 1997;67(3):345-66. doi: 10.2307/2963459.

8.   De Cáceres M, Legendre P. Associations between species and groups of sites: indices and statistical inference. Ecology. 2009;90(12):3566-74. doi: 10.1890/08-1823.1.

9.   Jackson DA. Compositional data in community ecology: The paradigm or peril of proportions? Ecology. 1997;78(3):929-40. doi: 10.1890/0012-9658(1997)078[0929:CDICET]2.0.CO;2.

10.   Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2011;8(9). doi: 10.1371/journal.pcbi.1002687.

11.   McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecol Lett. 2007;10(10):995-1015. doi: 10.1111/j.1461-0248.2007.01094.x.

12.   Volkov I, Banavar JR, Hubbell SP, Maritan A. Patterns of relative species abundance in rainforests and coral reefs. Nature. 2007;450(7166):45-9. doi: 10.1038/nature06197.

13.   Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA. 2002;99(16):10494-9. doi: 10.1073/pnas.142680199.

14.   Grundy PM. The expected frequencies in a sample of an animal population in which the abundances of species are log-normally distributed. Part I. Biometrika. 1951. doi: 10.2307/2332589.

15.   Bulmer MG. On fitting the Poisson lognormal distribution to species-abundance data. Biometrics. 1974;30. doi: 10.2307/2529621.

16.   Williamson M, J. Gaston K. The lognormal distribution is not an appropriate null hypothesis for the species–abundance distribution. J Anim Ecol. 2005;74(3):409-22. doi: 10.1111/j.1365-2656.2005.00936.x.

17.   Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol. 2012;8(10):e1002743. doi: 10.1371/journal.pcbi.1002743.

18.   Engen S, Lande R, Walla T, DeVries PJ. Analyzing spatial structure of communities using the two-dimensional poisson lognormal species abundance model. Am Nat. 2002;160(1):60-73. doi: 10.1086/340612.

19.   Bastidas-Oyanedel J-R, Mohd-Zaki Z, Pratt S, Steyer J-P, Batstone DJ. Development of membrane inlet mass spectrometry for examination of fermentation processes. Talanta. 2010;83(2):482-92. doi: 10.1016/j.talanta.2010.09.034.

20.   Preheim SP, Perrotta AR, Martin-Platero AM, Gupta A, Alm EJ. Distribution-based clustering: using ecology to refine the operational taxonomic unit. Appl Environ Microbiol. 2013;79(21):6593-603. doi: 10.1128/AEM.00342-13.

21.   Edgar RC, Flyvbjerg H. Error filtering, pair assembly, and error correction for next-generation sequencing reads. Bioinformatics. 2015. doi: 10.1093/bioinformatics/btv401.

22.   Techtmann SM, Fortney JL, Ayers KA, Joyner DC, Linley TD, Pfiffner SM, et al. The unique chemistry of Eastern Mediterranean water masses selects for distinct microbial communities by depth. PLoS ONE. 2015;10(3):e0120605. doi: 10.1371/journal.pone.0120605.

23.   Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7. doi: 10.1128/Aem.00062-07.

24. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. Bioinformatics. 2008;24(16):1757-64. doi: 10.1093/bioinformatics/btn322.

25. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 2011;21(3):494-504. doi: 10.1101/gr.112730.110.

26. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Meth. 2013;10(10):996-8. doi: 10.1038/nmeth.2604.

27. Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9(1):1-8. doi: 10.1186/1471-2105-9-386.

28. 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. Appl Environ Microbiol. 2006;72(7):5069-72. doi: 10.1128/aem.03006-05.

29. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1. doi: 10.1093/bioinformatics/btq461.

30. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26(2):266-7. doi: 10.1093/bioinformatics/btp636.

31. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490. doi: 10.1371/journal.pone.0009490.

32.   Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289-90. doi: 10.1093/bioinformatics/btg412.

33.   Human Microbiome Project Consortium. A framework for human microbiome research. Nature. 2012;486(7402):215-21. doi: 10.1038/nature11209.

34.   Head IM, Jones DM, Röling WF. Marine microorganisms make a meal of oil. Nat Rev Microbiol. 2006;4(3):173-82. doi: 10.1038/nrmicro1348.

35.   Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, Canion A, et al. Hydrocarbon-degrading bacteria and the bacterial community response in Gulf of Mexico beach sands impacted by the Deepwater Horizon oil spill. Appl Environ Microbiol. 2011. doi: 10.1128/AEM.05402-11.

36.   Palleroni NJ, Pieper DH, Moore ERB. Microbiology of Hydrocarbon-Degrading Pseudomonas. In: Timmis KN, editor. Handbook of Hydrocarbon and Lipid Microbiology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 1787-98.

37.   Rivers AR, Sharma S, Tringe SG, Martin J, Joye SB, Moran MA. Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. ISME J. 2013;7(12):2315-29. doi: 10.1038/ismej.2013.129.

38.   Mishamandani S, Gutierrez T, Aitken MD. DNA-based stable isotope probing coupled with cultivation methods implicates Methylophaga in hydrocarbon degradation. Front Microbiol. 2014;5. doi: 10.3389/fmicb.2014.00076.

39.   Iwaki H, Takada K, Hasegawa Y. Maricurvus nonylphenolicus gen. nov., sp. nov., a nonylphenol-degrading bacterium isolated from seawater. FEMS Microbiol Lett. 2012;327(2):142-7. doi: 10.1111/j.1574-6968.2011.02471.x.

40. Lo N, Kim KH, Baek K, Jia B, Jeon CO. Aestuariicella hydrocarbonica gen. nov., sp. nov., an aliphatic hydrocarbon-degrading bacterium isolated from a sea tidal flat. Int J Syst Evol Microbiol. 2015;65(6):1935-40. doi: 10.1099/ijs.0.000199.

41. Al-Awadhi H, Dashti N, Khanafer M, Al-Mailem D, Ali N, Radwan S. Bias problems in culture-independent analysis of environmental bacterial communities: a representative study on hydrocarbonoclastic bacteria. SpringerPlus. 2013;2:369. doi: 10.1186/2193-1801-2-369.

42. Jung S, Park J, Kown O, Kang J-H, Shim W, Kim Y-O. Effects of crude oil on marine microbial communities in short term outdoor microcosms. J Microbiol. 2010;48(5):594-600. doi: 10.1007/s12275-010-0199-2.

43. Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. Bioinformatics. 2001;17(6):509-19. doi: 10.1093/bioinformatics/17.6.509.